Comparative Biochemical Investigations of Cellobiohydrolase (Cel7A) of Trichoderma reesei

Effects of Loop Structure and Functions

Røjel, Nanna Sandager

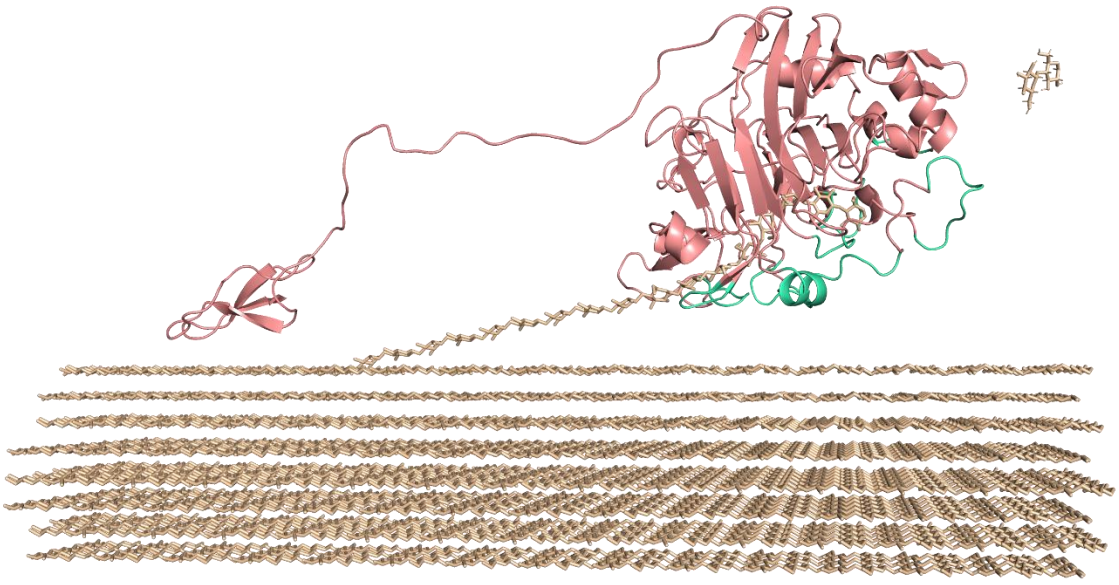
Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Røjel, N. S. (2019). *Comparative Biochemical Investigations of Cellobiohydrolase (Cel7A) of Trichoderma reesei: Effects of Loop Structure and Functions*. Roskilde Universitet.

Comparative Biochemical Investigations of Cellobiohydrolase (Cel7A) of *Trichoderma reesei:*

Effects of Loop Structure and Functions



May 2019

PhD thesis

Author: Nanna Røjel

Supervisor: Prof. Peter Westh

Co-supervisor: Science Director Kim Borch

This thesis has been submitted to the Doctoral School of Science and Environment

RUC

Summary

Utilization of fossil resources has revolutionized the modern society, however large quantities of CO_2 has been emitted and still is as a consequence of how we live. The growing amount of greenhouse gasses leads to global temperature increase, which has already led to pronounced impacts on the climate. Therefore, conversion from fossil sources to renewable energy is crucial, if we want to protect ourselves and the Earth from immense consequences. One of many solutions to the pending challenge, is to replace fossil fuels with bioethanol produced preferable from lignocellulosic materials. Lignocellulose is recalcitrant and therefore difficult to degrade, but some organisms, such as *Trichoderma reesei*, are able to do so, by secretion of enzymatic cocktails. Some of these are called cellulases, and of them the cellobiohydrolase I (CBH I) also known as Cel7A, is the most abundant. This is also a significant component of the industrial cocktail, and as the enzymes constitute a considerable amount of production costs, engineering for optimization at industrial relevant temperatures is much needed.

This dissertation mainly revolves around the Cel7A of *T. reesei* (mainly referred to as TrCel7A) with focus on the properties of the loops and the kinetics of it. The enzyme has a characteristic fold with eight peripheral loops (A1-A4 and B1-B4) forming the roof of the catalytic binding tunnel. Spanning this are four almost completely conserved tryptophans. We developed a method, utilizing the intrinsic tryptophan fluorescence to measure the complexation between the enzyme and the substrate in the tunnel. As the enzyme gets filled with substrate, water is excluded, and the fluorescence signal increases, as water is a quencher of fluorescence. By singularly mutating the tryptophans to alanine, we found that removal of the tryptophans had surprisingly little influence on the rate of complexation.

We also investigated the loops of this enzyme. TrCel7A and the related endoglucanase (Cel7B) are structurally similar, except from four heavily truncated loops that might reason the functional disparity between the two enzymes. We deleted the loops unique to TrCel7A singularly, followed upon several kinetic investigations. The results showed that the B2 loop was the key determinant for the kinetic characteristics of the CBH/TrCel7A. These findings were in line with another study, where we further investigated the B2 loop of TrCel7A. By site saturation of

the N200 position, we found that deletion of the asparagine, or substitution to a small amino acid, likewise shifted the kinetic parameters to resemble those of an EG.

Finally, we also examined how TrCel7A was affected by pH and temperature in presence of insoluble substrate. We found that TrCel7A was unaffected to the surrounding pH from 2 - 5.5, if the activity was tested at a temperature where the enzyme was stable at the given pH. This was in contrast to pH-profiles made on the basis of soluble substrate hydrolysis, of which a classical bell shaped curve appeared. Furthermore, this was in contrast to pH-profiles of TrCel7B with the more open binding cleft. It is natural to think that the accessibility to the catalytic binding site has an influence on this difference, and that this is governed by the truncated loop structure.

Together, this work contributes to the understanding of the molecular mechanism of this industrially relevant enzyme, TrCel7A. We elucidated enzyme-substrate complexation rates, stability, pH effects and kinetics of several of the important loops as well as identified the key loop for TrCel7A properties. These findings may be applicable for future enzyme engineering and improvement of the industrially process for bioethanol production.

Resume

Udnyttelse af fossile ressourcer har revolutioneret det moderne samfund, men store mængder CO_2 er blevet udledt og bliver det stadig som følge af, hvordan vi lever. Den stigende mængde drivhusgasser fører til global temperaturforøgelse, som allerede har medført markante konsekvenser for klimaet. Derfor er omlægning fra fossile kilder til vedvarende energi afgørende, hvis vi vil sikre os selv og Jorden fra endnu ukendte konsekvenser. En af mange løsninger på den igangværende udfordring er, at erstatte fossile brændstoffer med bioethanol fremstillet fortrinsvis af lignocelluloseholdige materialer. Lignocellulose er et genstridigt materiale og er derfor svært at nedbryde. Nogle organismer er dog i stand til at nedbryde lignocellulose, som for eksempel svampen *Trichoderma reesei*, der sekrerer en enzym-cocktail, som blandt andet indeholder enzymer, der nedbyrder cellulose, kaldet cellulaser. Af disse enzymer er cellobiohydrolase I (CBHI), også kendt som TrCel7A, den mest forekommende. Dette enzym er også en vigtig del af den industrielle cocktail til fremstilling af bioethanol. Da omkostningerne til enzymer udgør en betydelig del af produktionsomkostningerne, er det nødvendigt at designe og udvikle enzymerne til at virke optimalt ved industrielt relevante betingelser.

Denne afhandling handler hovedsagelig om Cel7A fra *T. reesei* (primært betegnet som TrCel7A) med fokus på egenskaberne af loop-strukturerne og kinetikken, der beskriver den. TrCel7A har en karakteristisk struktur med otte loops (A1-A4 og B1-B4), der danner et tag rundt om en katalytisk bindings-tunnel. I tunnelen findes fire næsten helt konserverede tryptofaner, der er spredt ud over tunnelen. Vi udviklede en ny metode, der udnytter den iboende tryptofan-fluorescens, til at måle kompleks-dannelsen mellem enzymet og substratet i tunnelen. Når enzymet fyldes med substrat, bliver vandet fortrængt, og fluorescencesignalet stiger, da vand dæmper signalet. Ved mutation af enkeltstående tryptofaner til alanin, fandt vi at tryptofanerne havde overraskende lidt påvirkning på hastigheden hvorved kompleks mellem substrat og enzym dannes.

Gennem omfattende studier undersøgte vi funktionen af flere af TrCel7As loops. Ved en sammenligning af TrCel7A med dens nære slægtning, endoglukanasen (EG) TrCel7B, træder strukturforskelle hurtigt frem, da fire af loopsne i TrCel7B er betragteligt afkortede i forhold til de tilsvarende loops i TrCel7A. Disse strukturelle forskelle er formentlig årsagen til den funktionelle forskel mellem enzymerne, som

ellers deler stor sekvenslighed. Enkeltvis fjernede vi de fire loops, som kun optræder i TrCel7A og fandt, at B2-loopet var en afgørende faktor for de kinetiske egenskaber der adskiller TrCel7A fra TrCel7B. Ved at udskifte N200-positionen i B2-loopet af TrCel7A med alle andre aminosyrer, fandt vi ud af, at når vi fjernede animosyreren (asparagin) eller skiftede den ud med en relativt lille aminosyre, ændredes de kinetiske parametre mod en EG.

Afslutningsvis undersøgte vi også dybtgående, hvordan pH og temperatur påvirker aktiviteten af TrCel7A på uopløseligt substrat. Vi fandt ud af, at aktiviteten af TrCel7A var uanfægtet af den omgivende buffers pH fra 2 - 5,5, såfremt at enzymet var stabilt ved den givne temperatur og pH. Den observation var i modsætning til pH-profiler lavet på basis af opløseligt substrat, hvoraf en klassisk klokkeformet kurve dukkede op. pH profiler for TrCel7B, som har et mere åbent bindingssite end TrCel7A, var ligeledes klokkeformede, og det er nærliggende at tænke, at tilgængeligheden til det katalytiske bindingssite har indflydelse på de forskellige pH profiler.

Tilsammen bidrager dette arbejde til forståelsen af molekylære mekanismer i dette industrielt relevante enzym. Vi belyste hastigheden af kompleksdannelse mellem enzym og substrat, stabilitet, pH-effekter og kinetikken af flere loop-varianter, såvel som identificerede det loop, som er vigtigst for de karakteristiske egenskaber der adskiller TrCel7A fra TrCel7B. Disse resultater kan forhåbentlig bidrage til fremtidige forbedringer af enzymer til den industrielle bioethanol-produktion.

Preface

This doctoral thesis is part of the requirements for a PhD degree at Department of Science and Environment at Roskilde University, Denmark. The work presented here was performed between June 2016 and May 2019 by supervision of Professor Peter Westh. The project was funded by Innovation Fund Denmark (no. 5150-00020B, granted to Peter Westh), Roskilde University and Novozymes A/S for a collaboration between these. The work during my final year was done at Department of Bioengineering at the Technical University of Denmark (DTU).

The work performed during the last three years, has been a part of a research project named TEMPEN (Temperature activation of enzymes for biomass degradation). In this project, the main purpose was to engineer better performing cellulases at industrially relevant temperatures, through an in-depth understanding of the molecular mechanisms of these enzymes. The first part remains confidential, for which reason this thesis builds on the latter. The dissertation is intended to give an entry to the appended publications, and therefore only selected details and results will be provided here.

In the light of the 2018 Nobel Prize in Chemistry recently given to Francis Arnold for her work on directed enzyme evolution, this work is highly topical. This, together with her work on cellulases and the current global climate situation, has really emphasized the importance and the potential of modeling and forming enzymes to the use of specific purposes. Hopefully, this work can benefit to a better understanding and development of cellulases.

"*Enzymes are by far the world's best chemists*"
- Frances H. Arnold. Professor at Caltech

Ørsted lecture, DTU May 2019:
"Innovation by Evolution: Bringing new chemistry to life"

Guide to the reader

This thesis mainly revolves around cellobiohydrolase TrCel7A with some references to the structurally similar endoglucanase, TrCel7B. The aim of this dissertation is to give an introduction to selected results based on the appended papers. Therefore, the reader is referred to the aforementioned papers for a comprehensive description of the work. The appended papers will be referred to by the roman numbers and marked by bold in the text (**paper I**, **II**, **III** etc.). In addition, the thesis has been supplemented with some unpublished results that are not a part of any papers, and this will be stated in the text.

The thesis is divided into four chapters. Chapter 1, *Introduction*, will provide a brief introduction to the background of the motivation, substrate and the enzyme. In the following chapters selected findings of the work done the past three years will be presented. Chapter 2, *Cellulase kinetics*, will briefly introduce cellulase kinetics and results from **paper III**. Chapter 3, *Exploring the properties of TrCel7A loops*, will revolve around the properties of loops unique to TrCel7A, and discuss the key determinant for cellobiohydrolase activity, **paper I**. Further, this chapter will summarize kinetic effects of two loops upon different engineering strategies, **paper IV**. Lastly, chapter 4, *pH effects on cellulases*, will treat the topic of pH and temperature relations during hydrolysis on both insoluble and soluble substrates, **paper II**.

Disclaimer

A minor part of this thesis is based on unpublished results. However, the enzymes and methods used to obtain the data are equivalent to methods of the appended papers.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Professor Peter Westh for being a great inspiration as well as for showing his deep love for science and enthusiasm for new results. I am grateful to have had have such direct access to your supervision, and your always objective and realistic view on every subject. Thank you for giving me the opportunity to guest DTU during the final year and for always wanting the best for me and the group. Thanks also to Science Director Kim Borch for technical views and lots of encouragement and making room for both celebration and failure. As you say "*perfect is the enemy of good*".

I also wish to thank the group; Stefan Jarl Christensen, Jeppe Kari, Silke Flindt Badino, Trine Holst Sørensen, Corinna Schiano de Cola, Ana Mafalda de Almeida Cavaleiro, Malene Billeskov Keller, Stine Fredslund Hansen, Michael Skovbo Windahl, Nanna Rolsted Sørensen, Radina Tokin, Kay Schaller, Bartlomiej Kolaczkowski and Gustavo Avelar Molina. It has been a pleasure for me to go to work with fantastic colleagues, and many of you have become my very good friends. You guys have lowered the activation energy for getting up in the morning and go to work – especially on the though days. This group has synergized and I have learned so much from you. Special thanks goes to my office mate, Stefan, for insightful discussions about science and private life, and to Jeppe for your endless stream of ideas, your love to science and eagerness to see new results and your patience. The same goes for Silke and Trine for your optimism and always good mood and just for making everything easier. Thank you for that!

Thank you Kenneth Jensen, Science Manager at Novozymes, to undertake and support the risky project of swapping loops between two enzymes and to make science playful and showing me how to venture into projects. Moreover, to notify about the opening of this PhD position and to encourage me to apply for it. Thank you Ria Jacobsson, Research associate, for engaging in all my projects, teaching me tips and tricks in the lab while cloning complicated constructions and for endless support to me and my work. Thank you Ulla Rosenberg and Camilla Hindborg Christensen and all the technicians at Novozymes for always helping out and being the bridge between the university and the Novozymes lab. Additionally, I would also like to thank Preben Morth for demonstration of the fluorospectrometer, ideas, excitement and

good collaboration and Camilla Graversen for good collaboration and your many hours in the lab.

Finally, I would like to express my gratitude to those closest to me. First, my close friends Eva, Caroline, Malene, Line, the Biotek'10-group and many more who have supported me on this journey through the PhD and for innumerable talks about nothing and everything. Lastly, thank you to my brother, Henrik, and my sweet mom and dad, Kirsten and Morten for always believing in me, having my back and cheering for me.

Nanna Røjel, May 2019

Publications

*Note that authors marked with asterisk contributed equally

I. **Systematic deletions in the cellobiohydrolase (CBH) Cel7A from the fungus *Trichoderma reesei* reveal flexible loops critical for CBH activity**
Corinna Schiano di Cola*, Nanna Røjel*, Kenneth Jensen, Jeppe Kari, Trine Holst Sørensen, Kim Borch and Peter Westh. *Journal of Biological Chemistry* (2019) Feb; 294(6) 1807–1815. doi: 10.1074/jbc.RA118.006699

II. **pH-profiles of cellulases depend on the architecture of the binding region**
Nanna Røjel, Jeppe Kari, Trine Holst Sørensen, Kim Borch & Peter Westh
Submitted to *Biotechnology and Bioengineering*, April 2019

III. **Strength and dynamics of substrate binding in the processive cellulase Cel7A: Transition state of complexation and roles of conserved Trp residues.**
Nanna Røjel, Jeppe Kari, Trine Holst Sørensen, Silke Flindt Badino, Preben Morth, Kay Schaller, Ana Mafalda Cavaleiro, Kim Borch and Peter Westh

IV. **Adaptive roles of a highly conserved asparagine residue in a loop of the processive cellulase Cel7A**
Trine Holst Sørensen, Silke Flindt Badino, Michael Skovbo Windahl, Nanna Røjel, Brett McBrayer, Kim Borch and Peter Westh
Submitted to *BBA Proteins and proteomics*, August 2019

V. **Thermoactivation of a cellobiohydrolase**
Peter Westh, Kim Borch, Trine Holst Sørensen, Radina Tokin, Jeppe Kari, Silke Badino, Ana Mafalda de Almeida Cavaleiro, Nanna Røjel, Stefan Jarl Christensen, Cynthia Segura Vesterager, Corinna Schiano-di-Cola. *Biotechnology and Bioengineering* (2018) Apr;115(4) 831-838. doi: 10.1002/bit.26513.

Abbreviations

CBH I	Cellobiohydrolase I, (Cel7A)
CBHII	Cellobiohydrolase II, (Cel6A)
CBM	Carbohydrate binding module
Cel6A	Cellobiohydrolase from glycoside hydrolase family 6
Cel7A	Cellobiohydrolase from glycoside hydrolase family 7
Cel7B	Endoglucanase from glycoside hydrolase family 7
DP	Degree of polymerization
EG	Endoglucanase
EGI	Endoglucanase I, (Cel7B)
GH7	Glycoside hydrolase family 7
MD	Molecular dynamics
MM	Michaelis-Menten
RAC	Regenerated amorphous cellulose
pNPL	para-Nitrophenyl-β-D-lactopyranoside
wt	wild type
DSF	Differential Scanning Fluorimetry
DSC	Differential Scanning Calorimetry
BG	β-glucosidase
CAZy	Carbohydrate active enzymes
PAHBAH	*para*-hydroxybenzoic acid hydrazide
pNPL	*para*-Nitrophenyl β-D-lactopyranoside
TEMPEN	Temperature activation of enzymes for biomass degradation
T. reesei,	*Trichoderma reesei*

Kinetic parameters and rate constants

${}^{\text{conv}}\text{V}_{\text{max}}$	Conventional maximum rate (saturated enzyme)
${}^{\text{conv}}\text{K}_{\text{M}}$	Conventional Michaelis constant
${}^{\text{inv}}\text{V}_{\text{max}}$	Inverse maximum rate (saturated substrate)
${}^{\text{inv}}\text{K}_{\text{M}}$	Inverse Michaelis constant
${}_{p}k_{on}$	Rate constant of association (processive enzymes)
${}_{p}k_{off}$	Rate constant of dissociation (processive enzymes)
${}_{p}k_{hyd}$	Rate constant for hydrolysis (processive enzymes)
${}_{p}k_{cat}$	Rate constant for catalysis (processive enzymes)
n	number of hydrolysis in a processive run
Γ_{max}	Maximum adsorption capacity
Γ_{attack}	Number of attack sites per gram substrate
K_{d}	Dissociation constant
K_{p}	Partitioning coefficient

Table of contents

1 Introduction

Utilization of fossil resources such as gas and oil, has throughout the 20th century revolutionized the foundation for modern society. Through refinery of these resources into energy and chemicals the building blocks to transportation, fertilizers for agriculture, materials such as plastics and textiles in addition to pharmaceuticals etc., our standard of living has been strongly improved. However, this way of living has an impact on the climate. The emission of CO_2 and other greenhouse gasses has increased during this period concurrently with the global temperature, and how that will further affect the Earth remains an unanswered question. In order to reach or stay below the aim from the Paris Agreement from COP21 (2015) of a global temperature increase of 2 °C, we need not only to limit the CO_2 emissions, we also need to capture CO_2 (IPCC, 2018). This altogether puts an emphasis on why we so strongly need to find ways to replace fossil resources with "clean" renewable resources. Political focus on sustainability such as UN Sustainable Development Goals (SDG) (www.sustainabledevelopment.un.org) and promotion of bioeconomy, show that there are several initiatives ongoing.

Lignocellulosic plants, a sustainable and CO_2 neutral material, are widely recognized for their potential as alternative to fossil fuels by deconstruction into fermentable sugars, followed by conversion into fuels and chemicals (Himmel, Ruth, & Wyman, 1999; Jørgensen, Kristensen, & Felby, 2007). In nature, the degradation of plants is a slow process. Plants are made to be able to withstand the danger of being degraded by competitors such as bacteria and fungi. The stem is produced to elevate the plant and is meant to be difficult to degrade by its competitors. This makes second generation bioethanol expensive to produce in contrast to first generation bioethanol where edible feedstocks are used as the energy source (Carriquiry, Du, & Timilsina, 2011). The advantage of second generation fuels is that waste and bi-products are used rather than potential human food such as wheat, corn, sugarcanes and sugar beets. This is much preferred as these resources are limited some places on Earth.

The overall process of second generation bioethanol production is to degrade pretreated feedstocks such as corn stover, wheat straw, bagasse or other saccharide-containing energy sources to small fermentable saccharides. The decomposition can be accelerated by enzymes from microorganisms like bacteria or fungi that are able to depolymerize lignocellulosic biomass into soluble sugars, which can later be fermented into ethanol by yeast. This part of the industrial process can be either

separate hydrolysis and fermentation (SHF) or simultaneous saccharification and fermentation (SSF) (Jørgensen et al., 2007). The ethanol can then be distilled and added as a supplement to gasoline or, by only minor modifications of normal cars, be the only energy source (**Figure 1**).

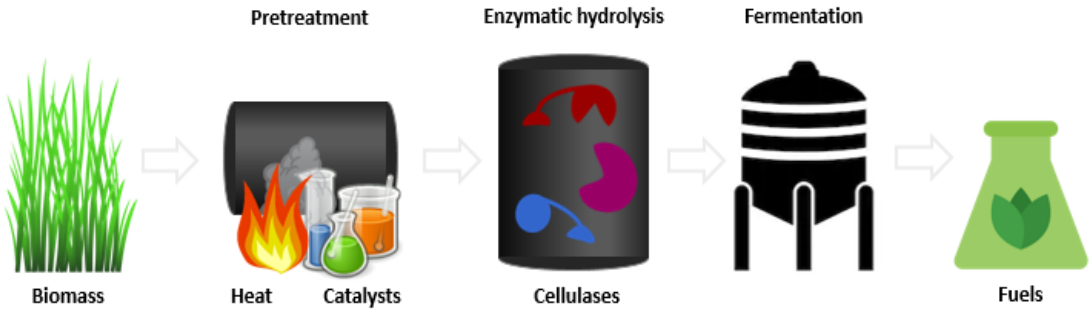


Figure 1: Overview of a simplifyed second generation bioethanol production.

The challenge is to make the overall bioethanol production profitable compared to fossil sources in order to implement this solution on a global scale. Many processes can be optimized, and the price of enzymes is one of them, as this make up a significant part of production costs (about 10-20 %) (Gregg, Boussaid, & Saddler, 1998; Wingren, Galbe, & Zacchi, 2003). The focus of this thesis revolves around depolymerization of cellulose and basic understanding of a central cellulase, *Trichoderma reesei* cellobiohydrolase I, which can help elucidate functions and mechanisms for further engineering of a more efficient enzyme.

1.1 Substrates

In order to investigate and improve a system, one has to understand the basic components to build a foundation of understanding. A key aspect of this system, is understanding the interaction between the enzyme and its substrate. As cellulases are in solution and its substrate, cellulose, is insoluble, the interaction between these happens in the solid-liquid interface. A central part of this dissertation revolves around the interplay between these, and in the following section some fundamentals of this simple but yet complex substrate, cellulose, will be reviewed.

1.1.1 Lignocellulose

The lignocellulosic rich cell wall of higher plants, also known as plant biomass, is the most abundant source to fermentable carbohydrates on Earth (Himmel et al., 1999; Pauly & Keegstra, 2008). This rich source of energy includes huge amounts of agricultural and forestry waste products, with an estimated yearly cellulose production of 10^{12} tons (Klemm, Heublein, Fink, & Bohn, 2005). During photosynthesis, solar energy is stored as chemical energy in plants. In the cell wall, most of this energy is stored as cellulose, hemicellulose and lignin, which constitute around 35-50%, 25-35% and 15-30% of the dry matter content, respectively (Jørgensen et al., 2007). Not all plant cell walls are composed similarly, as the composition vary between species, season of growth and production as well as other conditions (Pauly & Keegstra, 2008). To illustrate how these three components play together in the cell wall, one could imagine reinforced concrete. The cellulose would be the steel, reinforcing the cell wall with parallel structures of microfibrils, whereas the hemicellulose works as cables between the cellulose, providing transverse stabilization. Lastly the lignin cement it all together filling the gabs (Davison, Parks, Davis, & Donohoe, 2013).

The nature of the lignocellulosic composite matrix complicates the study of fundamental research. Consequently, many cellulase studies, like other enzymes, are performed using a simplified system, and in this case with pure cellulose substrates. This helps isolate the study of enzymes acting on one specific substrate before investigating more complex situations.

1.1.2 Cellulose

Cellulose is the most abundant biological macromolecule on Earth, and constitute an essential structural cell wall component of green plants but it is also present in some bacteria and algae (Morgan, Strumillo, & Zimmer, 2013). The glucoside bonds connecting the units, are extremely stable and abundant linkages in Nature with a half-life of spontaneous hydrolysis of millions of years (Wolfenden, Lu, & Young, 1998). This makes the cell wall strong and recalcitrant, and is the main challenge for decomposition of this material (Pauly & Keegstra, 2008).

In the plant cell wall, the cellulose chains are synthesized by large membrane bound multi-protein complexes called “rosettes”. They contain several cellulose synthase

subunits that elongates the polymer from the non-reducing end (Desprez et al., 2007; Lai-Kee-Him et al., 2002). These produce cellulose chains that are condensed into microfibrils (of approximately 24-36 chains) and further into fibers (Payne et al., 2015), see Figure 2.

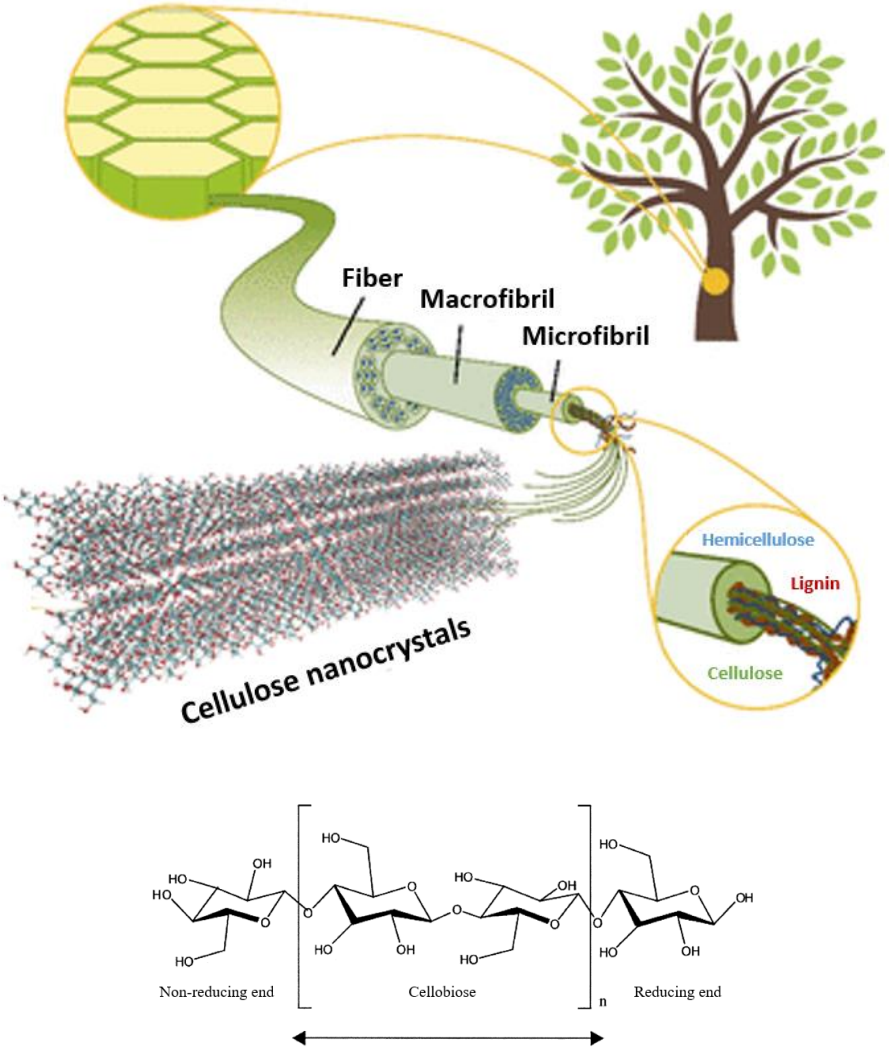


Figure 2: Schematic drawing of lignocellulose and its major components (cellulose, hemicellulose and lignin) down to the structure of cellulose nanocrystals and a single chain. Figure modified from (Martin-Martinez, 2018) and (Zhang & Lynd, 2004).

Cellulose is an insoluble unbranched polymer of glucose linked by β-1,4-glycosidic bonds with a degree of polymerization (DP) that ranges from hundreds and up to several thousands of glucoside units (Duchesne & Larson, 1989). The cellulose chain

has a reducing end able to open and form an aldehyde, while the opposite end cannot and is termed the non-reducing end (Figure 2). Chemically, the structure of cellulose is similar to that of starch, but as every second glucose unit is rotated 180°, the repeating units becomes cellobiose that is linked with β-linkages, contrary to α-linkages of starch. The alternating orientation of the pyranose rings results in an extended linear conformation that forms microfibrils with other chains facilitated by multiple inter- and intramolecular hydrogen bonds. In addition, hydrophobic stacking interactions between the glucose units also contributes to holding the chains in place and form a strong crystalline matrix of cellulose (O'sullivan, 1997). Altogether, these interactions make cellulose a very stable recalcitrant material to both chemical and enzymatic degradation and perfect as a building material in the plant and in buildings.

Fibrils of cellulose consists of both crystalline and amorphous parts and the ratio varies between cellulose sources (O'sullivan, 1997). The amorphous regions probably arise due to the twist of the cellulose chain which causes tension in the crystal that needs to be released in the more loosely packed amorphous regions (Matthews et al., 2006).

1.1.3 Model substrates

In this thesis, work has been performed on pure cellulose substrates. This has mainly been Avicel, but also Regenerated Amorphous Cellulose (RAC) and Bacterial MicroCrystaline Cellulose (BMCC) have been used.

Avicel is the most commonly used cellulose model substrate for cellulase investigations. Avicel is microcrystalline cellulose from American Viscose Company, where it has its name from. This substrate is made from wood chips, and the material consists of almost equal parts of crystalline and amorphous cellulose. The crystallinity index (CrI) is a parameter for crystallinity, which can be measured in different ways and for Avicel this is 0.5-0.6 (Hall, Bansal, Lee, Realff, & Bommarius, 2011; Reier, 2000; Zhang & Lynd, 2004). Amorphous cellulose, such as RAC, can be produced from Avicel by phosphoric acid treatment and has practically no crystalline parts left, resulting in a CrI of 0-0.04 (Zhang, Cui, Lynd, & Kuang, 2006). This substrate

is similar to Phosphoric Acid Swollen Cellulose (PASC), which is also often used as an amorphous substrate. In the other end of the "crystallinity spectra" is BMCC which is almost solely crystalline (CrI between 0.76 and 0.95). This is prepared by acid hydrolysis of the amorphous regions from Bacterial Cellulose (BC) (Väljamäe, Sild, Nutt, Pettersson, & Johansson, 1999), which leaves behind a mostly crystalline product.

In cellulase studies, substrate analogs are also used. They are often small synthesized molecules that upon enzymatic degradation releases an easily detectable molecule such as a chromophore or a fluorophore. Such analog could be para-nitrophenyl substrates (pNP) or para-methylumbelliferyl (MU) which for CBH studies often is substituted with lactopyranoside (Figure 3). Small substrate analogs are not always good alternatives for insoluble substrates when measuring activity and properties as will be discussed in **paper II**.

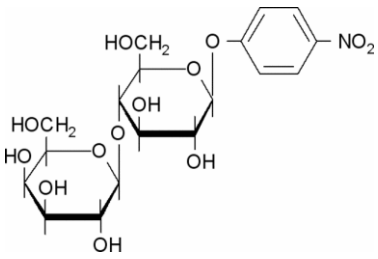


Figure 3. 4-Nitrophenyl β-D-lactopyranoside (pNPL). Figure from Sigma-Aldrich.com

1.2 Cellulases

Enzymes are biocatalysts that lower the energy barrier (activation free energy) of a reaction, and thereby increases the reaction rate. In many cases, the rate is enhanced millions and billions of times compared to the uncatalyzed reaction. In nature, cellulases are secreted by microorganisms in order to degrade cellulose to smaller units of mono-, di- or oligo-saccharides. The microorganism is then able to transport these over its membrane and utilize the energy of the sugars for its own growth. Cellulases from fungi are the best studied and most relevant for industrial utilization and in particularly the ones from the model organism *Trichoderma reesei* (Martinez et al., 2008).

The filamentous soft rot fungi, *T. reesei,* was for a short period denoted by its teleomorphic stage, *Hypocrea jecorina*, but the naming of this fungus has now returned to the more popular and well known, *T. reesei* since the International Biological Congress in Melbourne 2011 (henceforth used in this thesis) (Hawksworth, 2011). The name was assigned after the discovery during the Second World War, where a delegation of the US army was located in tropical and extremely humid region of the Salomon islands. A fungus was degrading the cotton canvas belonging to the army and from there it was isolated by E. T. Reese (Reese, 1976). *T. reesei* is mesophilic (Smits et al., 1998) encoding 200 glycoside hydrolases (GH) (Martinez et al., 2008) with a highly cellulolytic secretome including cellobiohydrolases (CBH) I and II and endoglucanases (EG) I, II, III and V. These are also referred to by their GH nomenclature as Cel7A, Cel6A, Cel7B, Cel5A, Cel12A and Cel45A, respectively (Henrissat, Teeri, & Warren, 1998; Martinez et al., 2008). Among these, the Cel7A and Cel6A are secreted in the largest amount, ~60% and ~20%, respectively, and this distribution is also reflected in the industrial cocktails (Rosgaard et al., 2007; Zhang & Lynd, 2004).

1.2.1 Nomenclature

The GH nomenclature were established by Henrissat et al. (1998) and the system is collected in the CAZY database (Carbohydrate Active enZYmes, www.cazy.org) (Lombard, Golaconda Ramulu, Drula, Coutinho, & Henrissat, 2013). The system is based on the enzyme's amino acid sequence and thereby indirectly also the fold. This is in contrast to the substrate specificity as used by the Enzyme Commission system (EC numbering) by International Union of Biochemistry and Molecular Biology (IUBMB). Using the substrate specificity is sometimes less convenient as several enzymes display catalytic activity towards more than one substrate and also because some function in more than one way (Henrissat et al., 1998). For instance, the distinction between cellobiohydrolases and endoglucanases is not that clear. While the CBH is primarily exoacting, working from the end of the cellulose strand, it has also been reported to be able to make endo-attacks (Silveira & Skaf, 2018; Ståhlberg, Johansson, & Pettersson, 1993) and therefore in principle could also be categorized as endo-active. Likewise, there are also cases where the distinction based on the structure is not sufficient underlining the necessity of both systems. The

nomenclature of the cellulases used in this work is outlined in Table 1, however the GH nomenclature is the preferred one.

Table 1. Different nomenclatures of cellulases used here, structural and functional.

Structural (GH)	Functional
Cel7A	CBHI
Cel7B	EGI
Cel6A	CBHII

1.2.2 Cellulases in synergy

In order to efficiently depolymerize cellulose, a cocktail of enzymes with different specificity needs to synergize, see Figure 4 (Kleman-Leyer, SiiKa-Aho, Teeri, & Kirk, 1996). The endoglucanases mainly work endo-actively on amorphous regions quickly decreasing the DP of the cellulose. It produces new reducing and non-reducing ends for the CBHI and the CBHII, respectively. While attached, these work processively, meaning that several bonds are hydrolyzed before disassociation from the cellulose. The product of the CBHs are predominantly cellobiose, which is cleaved to glucose units by the β-glucosidase (BG). As CBHs are inhibited, mostly by cellobiose, the presence of a BG helps the CBHs by decreasing this inhibition. In addition to the cellulases, is the relatively newly discovered oxidoreductases, also known as lytic polysaccharide monooxygenases (LPMOs) (Vaaje-Kolstad et al., 2010). They also contribute to the synergy by oxidizing glycosidic bonds leaving the substrate more susceptible to attack by other enzymes (Quinlan et al., 2011). After a cocktail of enzymes having worked their way through the cellulose, released glucose can now be taken up as feed by the organism that produced the enzymes.

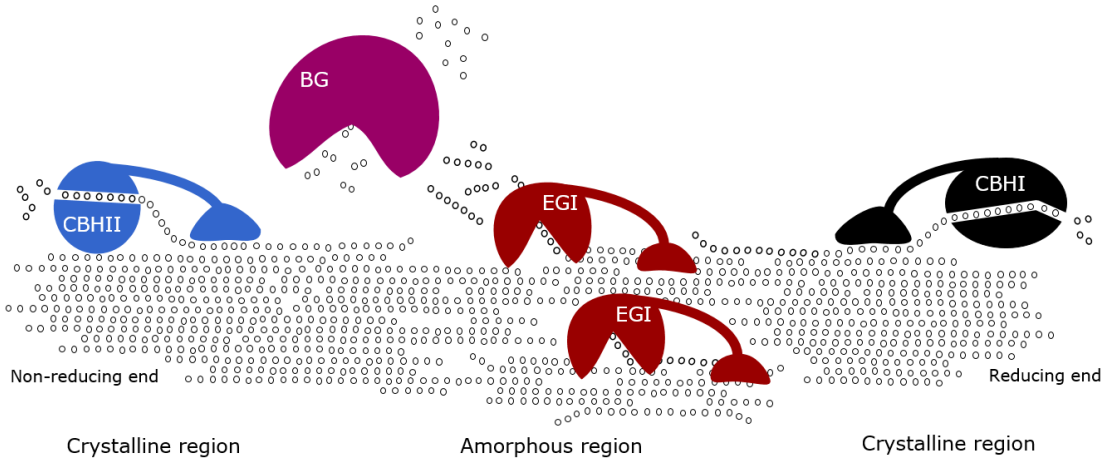


Figure 4: Illustration of synergism by cellulases. Cellobiohydrolase I, exo-acting from the reducing end, and II, exo-acting from the non-reducing end, (CBHI and CBHII), Endoglucanse I, endo-acting (EGI) and β-glucosidase (BG).

1.2.3 TrCel7A

TrCel7A is a cellobiohydrolase working exo-lytically from the reducing end of a cellulose polymer. After adsorption to the strand, it works processively by catalyzing several hydrolytic cleavages before desorbing again. Cel7A is comprised of a catalytic core domain (CD), a glycosylated linker region and a carbohydrate binding module (CBM). The CD has a β-jelly roll fold with peripheral loops stretching round a distinct catalytic binding area forming a tunnel. High resolution X-ray studies revealed that the binding tunnel extents ten pyranose binding sites (Divne, Ståhlberg, Teeri, & Jones, 1998) (Figure 5). Seven of these are 'minus' subsites where the cellulose strand enters the tunnel, while three 'plus' subsites are located in the expulsion area. The subsites -1 and +1 surrounds the location of the catalysis.

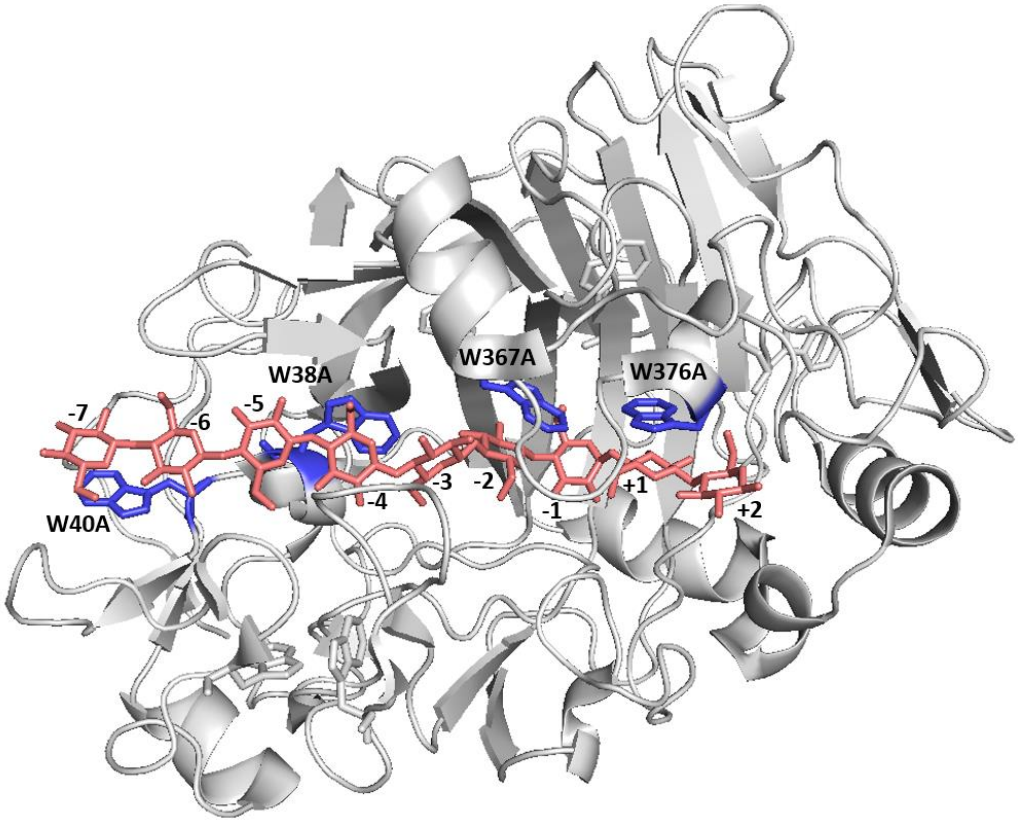


Figure 5: Cartoon of the catalytic core domain of TrCel7A in grey (PDB 4C4C (Knott et al., 2013)). The four tryptophans in the tunnel are highlighted in blue sticks. A cellononaose is shown in red sticks. Figure prepared with Pymol.

In the 50 Å long catalytic tunnel, four out of the total nine tryptophan residues are lining the binding tunnel (Figure 5). W40 is located in the entrance (-7 subsite), while W38 is positioned further in the middle (-4 subsite) and surrounding the catalytic residues are W367 and W376 (-2 and the +1 subsite, respectively) (Divne et al., 1998). The tryptophans contribute to binding of the substrate by stacking interactions with the glucosyl unit near it. These tryptophans are highly conserved between GH7s (Payne et al., 2015), which suggests that they are of vital importance. In the other very abundant GH7 enzyme, the TrCel7B share the same distribution of tryptophans except that the tryptophan at the position equivalent to W38 is a tyrosine (Kleywegt et al., 1997). Kari et al. (2014) mutated W38 in TrCel7A to an alanine and found that the enzyme had reduced substrate affinity and increased maximum specific rate. They proposed that the increased activity were a consequence of the lowered affinity to the substrate and thereby a faster dissociation rate. Experimental results (Igarashi et al., 2009; Nakamura et al., 2013) and MD

simulations (GhattyVenkataKrishna et al., 2013; C. B. Taylor et al., 2013) have reported high importance of W40 for the threating of crystalline cellulose. A comprehensive investigation of all the tryptophans has not previously been reported (Payne et al., 2015), but **paper III** elucidate the rate of threading (which we will return to in chapter II) and report the transition state for complexation as well as the dynamics of the tunnel tryptophans.

The loop nomenclature of TrCel7A used in this thesis has been suggested by Momeni et al. (2013). They divided the peripheral loops reaching out from each side, A and B, respectively (Figure 6). On each side, four loops extents towards each other forming a roof of the tunnel. These are named A1-A4 and B1-B4. The numbering follow the order of the loops of the tertiary structure, rather than the sequence. It has been suggested that the loops are responsible for the functional differences between CBH and EG. Meinke et al. (1995) removed one of the two loops of the related TrCel6A and found the endolytic activity of the mutant to become more EG-like. Likewise, deletion of the B3 loop of TrCel7A, resulted in reduced activity on crystalline cellulose in favor of increased activity on amorphous cellulose, hence, an increase in EG-like behavior (Von Ossowski et al., 2003). Several computational studies have supplemented the experimental work and extended our understanding of loop function and properties (Bu, Crowley, Himmel, & Beckham, 2013; Bu et al., 2012; Momeni et al., 2013; Silveira & Skaf, 2018; C. B. Taylor et al., 2013).

The sequential and structurally very similar TrCel7B (Penttilä, Lehtovaara, Nevalainen, Bhikhabhai, & Knowles, 1986), has in contrast to TrCel7A only four of the eight loops. This seems to be the major difference between the two, and it is natural to think that the loops unique to TrCel7A are responsible for the characteristics of the cellobiohydrolase. This idea has been investigated in **paper I**.

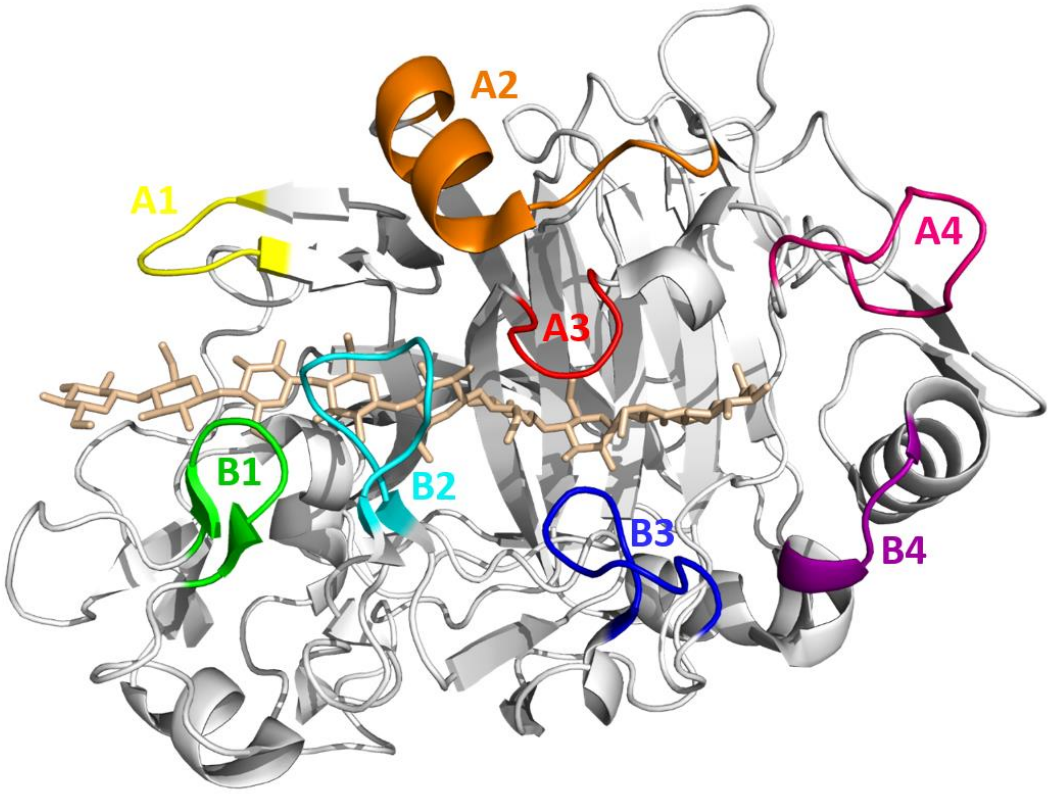


Figure 6: A cartoon of TrCel7A CD highlighting the loops. The cellononaose is shown in beige sticks and the loops are highlighted in colors with labels. PDB 4C4C (Knott et al., 2013). Figure made using Pymol.

The C-terminal end of the core domain is associated with a carbohydrate binding module belonging to family 1 (CBM1) through a glycosylated linker. As the name implies, this module has the function to increase the concentration of enzyme on the carbohydrate surface, because of its ability to bind efficiently to cellulose (Igarashi et al., 2009; Linder & Teeri, 1997; Payne et al., 2015; Várnai, Siika-aho, & Viikari, 2013). The three dimensional structure solved by Kraulis et al. (1989) using NMR, revealed a module formed as a wedge with a plane surface dominated by three tyrosines (marked in red, Figure 7) that interacts with the cellulose surface (Linder et al., 1995).

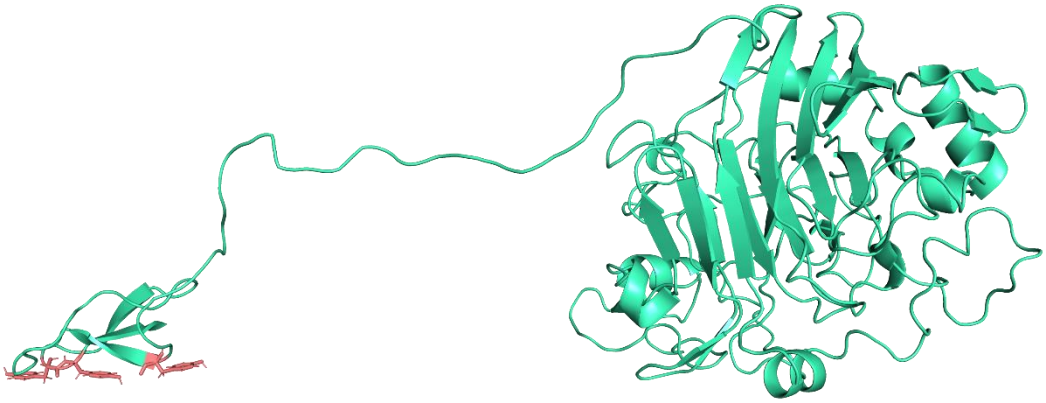


Figure 7: A suggestion of how the CD, linker and CBM are positioned compared to each other. The cartoon is based on 4C4C and 1CBH (Kraulis et al., 1989). Three tyrosines in the CBM are highlighted in red sticks. Figure prepared in Pymol by J. Kari.

The CBM has also been attributed the function of loosening the crystalline parts of the cellulose, and thereby making it more accessible resulting in higher product yields (Hall et al., 2011; Lemos, Teixeira, Domingues, Mota, & Gama, 2003). Furthermore, experimental investigations has showed that the CBM assists the enzyme in "finding" and sticking to the cellulose surface, and therefore is important during dilute substrate conditions. Contrary, it has been found that during high substrate concentrations, the enzyme can easily find a new place to attack, and the CBM can therefore be disadvantageous while keeping a too tight a grip on the cellulose (J. P. Olsen, Kari, Borch, & Westh, 2017; Sørensen et al., 2015; Várnai et al., 2013)

1.2.4 Catalytic mechanism of TrCel7A

Spontaneous hydrolysis of O-glycosides happens much slower than joints of other biological polymers (Wolfenden et al., 1998). With a half-life of a glycosidic bond spontaneous hydrolysis estimated to be 5 - 22 million years, an acceleration of the catalysis up to a couple of cuts per minute or several thousand per second, makes glycoside hydrolase some of the most proficient enzymes (Wolfenden et al., 1998; Zechel & Withers, 2000).

The catalytic mechanism of cellulases are either retaining or inverting. The retaining method preserves the anomeric conformation of the substrate using two steps, while the inverting method switch the anomeric conformation in a single step. Cel7A and

Cel6A are examples of enzymes with such catalytic mechanisms, respectively (Davies & Henrissat, 1995; Knowles, Lentovaara, Murray, & Sinnott, 1988; Rouvinen, Bergfors, Teeri, Knowles, & Jones, 1990).

Throughout the glycoside hydrolases, a carboxyl pair constituting the catalytic residues, remains conserved (Zechel & Withers, 2000). In TrCel7A these residues, Glu212, Asp214 and Glu217, are closely located. In fact, they are situated on the same β-sheet surrounding the scissile bond of the cellulose strand (Figure 8). Ståhlberg et al. (1996) found that E212 is the most important of the catalytic residues as only 1/2000 of activity is retained after substituting the glutamate by a glutamine. Using high speed atomic force microscopy Igarashi et al. (2009) showed how the enzyme processively moved along the cellulose strand. They also showed, retrospectively not surprisingly, that catalytic activity is a prerequisite for processivity, as the E212Q mutant could attach, but not move down the crystalline cellulose string.

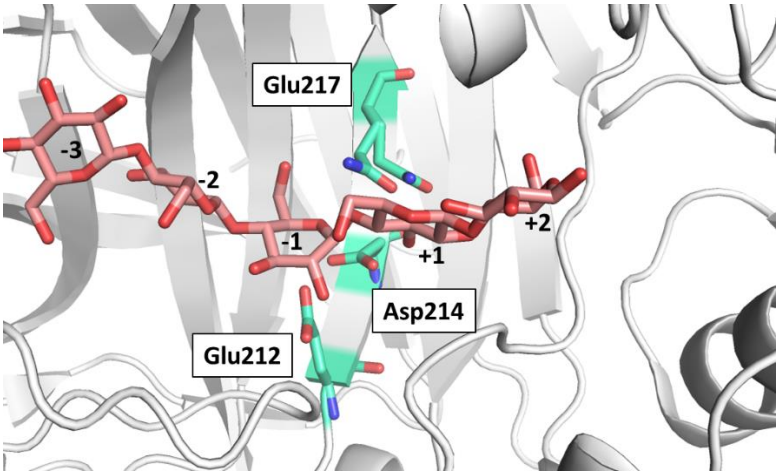


Figure 8: The three catalytic residues are all located on one beta-sheet surrounding the scissile bond between -1 and +1. Figure prepared with Pymol.

The mechanism of TrCel7A is a double displacement reaction, which starts by a proton being transferred from the Glu217 (the catalytic acid/base) followed by an attack of the anomeric carbon by the nucleophile Glu212. This results in a covalent glycosyl-enzyme intermediate (Figure 9). Subsequently, a water molecule attacks the anomeric carbon resulting in bond cleavage of the intermediate and replacement of the bond back to Glu212. A proton transfer regenerates Glu217 and the enzyme is

ready to catalyze the next reaction (Bu et al., 2013; Payne et al., 2015). Asp214 also plays a role in the catalysis, perhaps attenuating the pK_a of Glu212 and keeping it in position via a hydrogen bond (Ståhlberg et al., 1996; Ståhlberg et al., 2001).

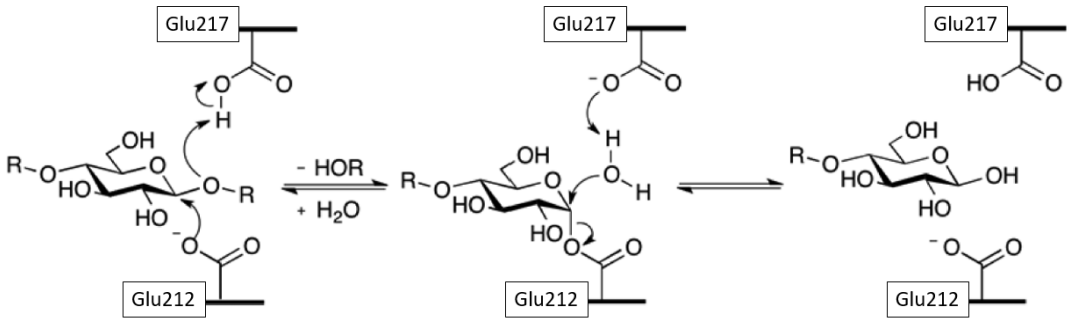


Figure 9: A schematic drawing of the catalytic retaining mechanism of TrCel7A. Figure modified from (Payne et al., 2015).

1.3 Protein engineering

In Nature, evolution is the driver for changes in biodiversity and for the characteristics of all kinds of biological aspects over successive generations. In the laboratory, with regards to proteins, we speed up the process and call it protein engineering. Protein engineering is the process of making changes in the amino acid sequence with the purpose of altering the properties and performance of a protein. In this specific case, we engineer cellulases in order to improve the overall hydrolytic performance at industrially relevant temperatures or to investigate the properties and bottlenecks of the enzymes. For this, we have mainly used rational protein design, a tool based on knowledge to direct mutations at specific locations in the enzyme. Alterations of the enzyme often start by simply studying the 3D structure of the enzyme. The designs are based on knowledge, which could be multiple alignments of one's favorite enzyme in addition to other enzymes that has the properties of interest, such as high performance at high temperature or thermostability. Rational design can also be based on experiments and interpretations to improve a single property as affinity (**paper V**) or by domain swapping (**paper I**) to try to move functions governed by a specific module of the enzyme. Another semi-rational approach of engineering could be to explore a carefully selected location by site saturation where one specific amino acids is replaced by all

other possible amino acids (**paper IV**). Finally, we also investigate the properties of conserved single amino acids with distinct functions (**paper III**).

Protein engineering is often an iterative process (Figure 10) where the benchmark enzyme is experimentally characterized to determine kinetic parameters and binding properties in addition to update of the current hypothesis. This is followed by analysis of additional modelling, identification of bottlenecks and identification of molecular origin to rate limiting steps of the catalytic cycle. This leads to redesign of the enzyme, perhaps with inspiration from bioinformatics or other well-performing variants that are cloned via site directed mutagenesis, fermented and purified. After this process, the iterative process is started all over again by repeating the characterization and validating the hypothesis.

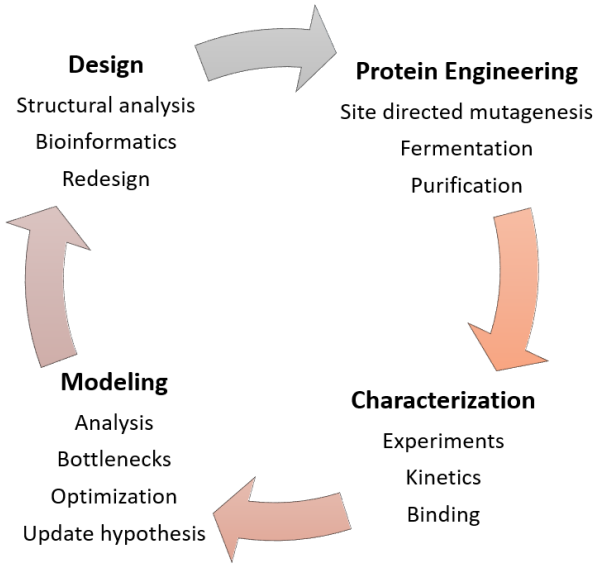


Figure 10: An iterative process for optimizing of enzymes.

1.3.1 Cellulases in other industries

Bioenergy is not the only field where cellulases are being used. They are also used as a component in detergents both for dishwashing and laundry. For laundry it is used to maintain and restore a 'new appearance' of clothes by cutting the fluffy lint that appear during usage and by maintaining clear colors (H. S. Olsen & Falholt, 1998). Cellulases are also used in the textile industry for similar purposes. In the pulp and

paper industry, cellulases are used for preparing and refining the fibers as well as helping de-ink the pulp (Bhat, 2000). These enzymes also have many applications in the food industry; for fruit and vegetable juice production, they are used for improvement of extraction, clarification and viscosity adjustment and amongst other applications, they are also used for both beer and wine production as well as olive oil extraction. Lastly, cellulases are added to feed for improvement of the nutritional quality (Bhat, 2000).

For the modern industrial enzymatic hydrolysis, many enzymes need optimization. This is important as enzymes are expensive and a considerable part of production cost (Jørgensen et al., 2007; Martinez et al., 2008; Wingren et al., 2003). Therefore, the processes need to be as efficient as possible in order to for example lower the bioethanol price and thereby also increase the incentive for the costumer to choose bioethanol over fossil fuels. Despite many known thermophile organisms, cellulases has not evolved to degrade cellulose at industrial relevant temperatures (ex. 50 °C or above), and because of this, we need to engineer enzymes optimized for this. In order to do so, one needs to understand the functions of the enzyme of interest and how they influence its kinetics and binding properties. We will examine these subjects in the next chapters.

2 Cellobiohydrolase kinetics

Enzyme kinetics is the study of the reactions catalyzed by the enzyme. It is an important tool to investigate the single steps of the mechanism, and therefore links the function and structure of a specific enzyme. However, the kinetics of enzymes working at the interface between the soluble and insoluble phase are poorly understood. Characterization of interfacial enzymes, such as cellulases, is in contrast to enzymes working solely in solution difficult as multiple phases of the enzyme catalysis exists. The complexity of cellulases is caused by for instance adsorption to the solid surface, threading, processivity, and desorption. Additionally, the heterogeneity of the substrate further adds to the complexity of the characterization.

Because of cellulose's nature as an unbranched polymer purely consisting of cellobiose units looks fairly trivial, it can be troublesome to work with kinetically. From an enzymes point of view, parameters like crystallinity, DP, and purity such as remaining of other components as lignin affect the enzymatic catalysis of the enzyme (Zhang & Lynd, 2004). These factors influence the accessibility for the enzyme to interact with the cellulose and find chain ends from which hydrolysis can be initiated (henceforth called attack sites) (Lynd, Weimer, Van Zyl, & Pretorius, 2002). Additionally, the insolubility of cellulose, makes it difficult to define the molar concentration of the substrate, as the total number of glycosidic bonds of the cellulose is not available to the enzyme. Even though we try, this will only be an estimate of the concentration. For a cellobiohydrolase, only the cellulose ends on the outside of the crystal are attackable. Only sites that are solvent exposed can be attacked by the enzyme. Therefore, many of the ends within a crystal are inaccessible. However, the number of attack sites is here estimated to be constant for a period, as new attack sites appear concurrently with those being removed.

In this chapter, kinetic tools for investigation of cellulases will briefly be introduced. These models originate from previous work from this research group (Cruys-Bagger, Elmerdahl, Praestgaard, Borch, & Westh, 2013; Kari, Andersen, Borch, & Westh, 2017; Praestgaard et al., 2011). Some main findings of kinetic investigations measured with real-time-fluorescence is described in the end of the chapter.

2.1 Classic enzyme kinetics

Before diving into specific cellulase kinetics, we will briefly recap some traditional enzyme kinetics. The latter builds on well-established knowledge, which is thoroughly described in text-books as (Berg, Tymoczko, & Stryer, 2012; Fersht, 1999; Nelson, Lehninger, & Cox, 2008). In traditional enzyme kinetics, we typically meet a scheme (Scheme 1) where a substrate, S, is converted into a product, P, which is facilitated by the catalyst, the enzyme, E, which is regenerated into its starting condition without being depleted. In this case k_1 is the association rate constant, k_{-1} is the dissociation rate constant, and k_2 is the rate constant for conversion from enzyme-substrate complex, ES, to product (the catalysis). This latter can also be termed the turnover number, reflecting the number of product produced per time per enzyme.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E + P \qquad (1)$$

If we consider a situation where the enzyme is conserved, $E_0 = E + ES$ (E_0 = initial enzyme concentration), and the substrate is in large excess compared to the enzyme, $S_0 \gg E_0$, then the conversion of substrate and the substrate concentration in the ES complex is negligible so $S \approx S_0$ (S_0 = initial substrate concentration). If we further apply the quasi steady-state approximation (QSSA), an assumption where concentrations for all intermediates in a period of time do not change, $\frac{d[ES]}{dt} \approx 0$, we can derive the famous Michaelis-Menten equation.

$$v = \frac{V_{max} \cdot S_0}{K_M + S_0} \qquad (2)$$

Here v is the initial rate of conversion, product per time, V_{max} is the maximum initial speed of conversion which is achieved when the enzymes is saturated with substrate, and K_M, the Michaelis-constant, which is the substrate concentration at half saturation (the rate equal to half of V_{max}). By plotting the rate versus the substrate concentrations, and fitting equation 2 to the experimental data, V_{max} and K_M can be derived. The two parameters are defined as below, and k_{cat} here reflects the slowest forward moving step.

$$K_M = \frac{k_2 + k_{-1}}{k_1} \text{ and } V_{max} = E_0 \cdot k_{cat}$$

Steady-state approximations can also be applied for the inverse situation, where the enzyme is in excess compared to the substrate. This is rarely relevant for homogeneous catalysis, as the substrate would be depleted instantly. Nevertheless, if

we consider this rare or hypothetical case where $E_0 \gg S_0$, the inverse Michaelis-Menten model can be derived, equation 3 (Bailey, 1989).

$$v = \frac{V_{max} \cdot E_0}{K_M + E_0} \qquad (3)$$

2.2 Processive cellulases

One thing that makes cellobiohydrolases different from most other enzymes, is that they work in a processive manner. This means that after association to the substrate, they stay attached while catalyzing several hydrolysis steps, before dissociation from the substrate. A model taking this into account was proposed by Praestgaard et al. (2011), a former member of this research group. A scheme visualizing this model is presented in Figure 11. Here the enzyme, E, associates with the cellulose, C, forming an EC_m complex, where the subscripted "m" reflects the degree of polymerization of cellobiose. The rate constant ${}_pk_{on}$ governs the rate of the EC_m complexation. Now, two scenarios can happen; either the enzyme dissociate from the substrate, governed by ${}_pk_{off}$, or catalysis of a glycosidic bond takes place, governed by ${}_pk_{hyd}$. The hydrolysis results in release of a cellobiose, and this process is repeated a number of times, which is designated by "n" that is the average number of processive cycles happening before the enzyme dissociates again.

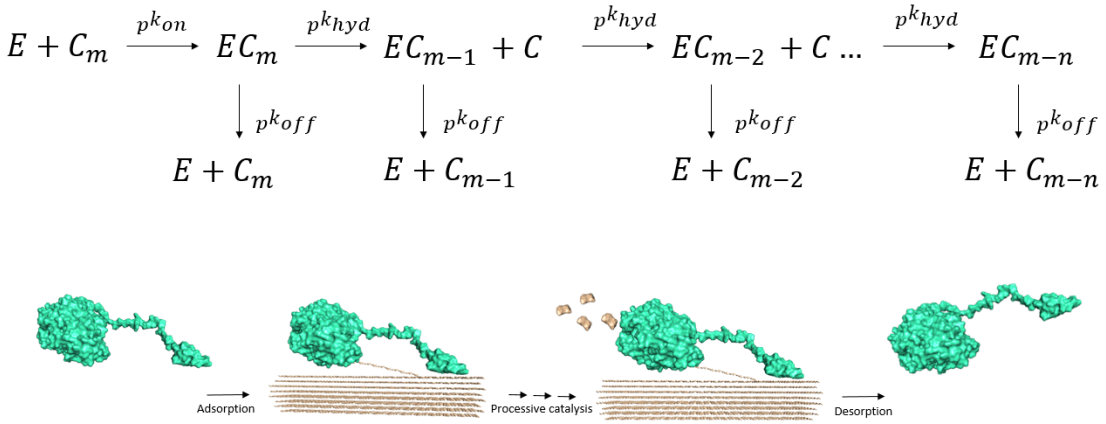


Figure 11. A simplified micro kinetics scheme of processive cellobiohydrolases (Praestgaard et al., 2011) (top), see text for further explanation, and an illustrative representation of the reaction (bottom). The reaction starts by adsorption of the enzyme (cyan) to the substrate (beige). The enzymes moves processively and catalyzes several cuts releasing products before it dissociate again.

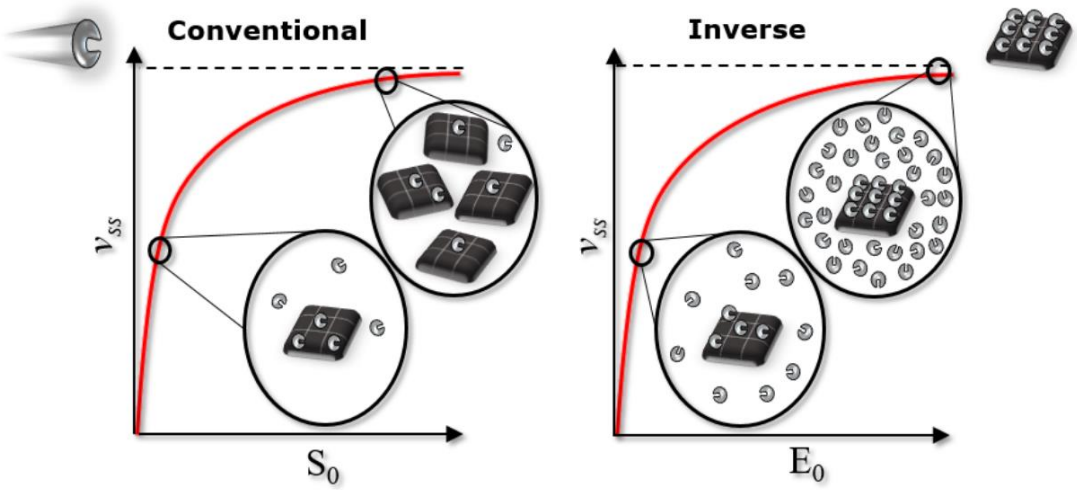
In order to determine ${}_pk_{on}$, ${}_pk_{hyd}$, and ${}_pk_{off}$, it is necessary to measure complicated pre-steady-state events. Due to the insolubility of the substrate, this is a quite troublesome matter for cellulases that requires specialized equipment, complicated analysis and fitting of long equations, however it has previously been investigated (Cruys-Bagger et al., 2012; J. P. Olsen et al., 2017). By using this special equipment on TrCel7A, a fast catalytic rate has been observed in the initial phase of the reaction, followed by a slowdown of the catalytic rate (Cruys-Bagger et al., 2012; Cruys-Bagger, Tatsumi, Ren, Borch, & Westh, 2013; Praestgaard et al., 2011). The molecular origin of such burst can be interpreted as a phase where the enzyme quickly associates to the substrate and the processive hydrolysis starts. Product are quickly released and the fast conversion rate is maintained, until the enzyme needs to dissociate from the substrate, causing the rate to decrease. Thus, it has been suggested that the dissociation rate is slower than the hydrolytic rate, and thereby the catalytic reaction of processive cellulases is dissociation-limited under some conditions (Cruys-Bagger et al., 2012; Kurašin & Väljamäe, 2011). Processivity is controlled by the substrate, and it has been suggested, that the enzyme eventually encounters steric irregularities on the substrate surface, known as obstacles, which blocks the movement of the enzyme (Igarashi et al., 2009; Jalak & Väljamäe, 2010; Praestgaard et al., 2011). Consequently, a concentration of non-productive enzyme builds-up in front of these obstacles. Since the dissociation is the slowest reaction step (e.i $k_{hyd} >> k_{off}$, Figure 11), this causes the observed slow-down in the rate (burst). As we will discuss later in chapter 3, a decreased affinity or a higher K_M will often lead to higher maximal rate, as the enzyme can escape the substrate easier and find a new sites to attack (Cruys-Bagger et al., 2012; Cruys-Bagger et al., 2013; Kari et al., 2014; Praestgaard et al., 2011).

Simple end-point measurements can effectively be used to approximate the kinetics of the steady-state phase subsequent to the burst, as further elaborated later in this chapter and exemplified several places in this work. These models require less laborious and troublesome experiments than the one described above, however the processivity is not the focus here, but instead some very applicable parameters can be derived from these models. Cruys-Bagger et al. (2013) refined and applied this

model to derive a steady-state theory for cellulases that will be presented in the following alongside with the recently proposed inverse steady-state mode.

2.3 Conventional and Inverse steady-state models

Cellulases display "double saturation" meaning that the rate of hydrolysis levels off at saturation of either the substrate or enzyme (Bezerra & Dias, 2004; Lynd et al., 2002). Therefore, we use two kinds of kinetic models; that we call the conventional Michealis-Menten model and the inverse Michealis-Menten model that will be described below. Using the conventional MM model, the enzyme is saturated with substrate, while the inverse MM approach require that the substrate is saturated with enzyme. Based on an steady-state approximation for the reaction intermediates, Cruys-Bagger et al. (2013) derived an equation for processive cellulases that resembles the MM-equation in equation 3. This will henceforth be called the conventional-MM equation, Figure 12. The same assumptions are fundamental for the inverse steady-state MM approach (Bailey, 1989; McLaren & Packer, 1970) recently applied on cellulases by Kari et al. (2017), Figure 12. The inverse method is possible for enzymes acting on cellulose. It is so, because we assume that not all attack sites are available initially, but keeps appearing as the outermost layer is being peeled of the cellulose crystal and because we do assume that the number of attack sites is conserved for a while. Important for both models is that the rate is measured before any considerable amount of substrate has been converted. The specified models are built on the classical methods, however V_{max} and K_M are differently defined as discussed in detail earlier (Cruys-Bagger et al., 2013) and shown in Figure 12.



$$^{conv}v_{ss} = \frac{^{conv}V_{max} \cdot S_0}{^{conv}K_M + S_0} \qquad ^{inv}v_{ss} = \frac{^{inv}V_{max} \cdot E_0}{^{inv}K_M + E_0}$$

$$^{conv}V_{max} = E_0 \cdot {}_pk_{cat} \qquad ^{inv}V_{max} = S_0 \cdot {}_pk_{cat} \cdot \Gamma_{attack}$$

$$^{conv}K_M \equiv \frac{{}_pk_{off}}{{}_pk_{on}} \qquad ^{inv}K_M = \frac{{}_pk_{off} \cdot {}_pk_{cat}}{{}_pk_{on}}$$

$$Assumption: S_0 + {}^{conv}K_M \gg E_0 \qquad Assumption: E_0 + {}^{inv}K_M \gg S_0$$

Figure 12: Illustrations of the conventional and inverse steady-state model (top) as well as central equations and definitions (bottom). Illustrations made by J. Kari, (Kari et al., 2017). The illustrations represents the steady-state rate (v_{ss}) as a function of S_0 at substrate excess (ConvMM) or E_0 at enzyme excess (InvMM). At convMM conditions the enzymes (grey) are saturated with substrate depicted as black squares representing cellulose ends the enzyme can attack. During invMM, the enzyme is in excess covering the all binding sites.

The **conventional Michaelis-Menten** model is applied for saturation, which means that the enzyme is saturated with substrate, reflecting the maximal catalytic rate of each enzyme, as principally all enzymes are in complex. The conventional maximum specific rate ($^{conv}V_{max}/E_0$) measured here is equivalent to normal turnover frequency ($_pk_{cat}$) which is defined earlier (Cruys-Bagger et al., 2013). Under **inverse**

Michaelis-Menten conditions, saturation means that essentially all attack sites are saturated with enzyme forming a complex and an excess of enzyme occur freely in the aqueous phase. Here, the inverse maximum specific rate ($^{inv}V_{max}/S_0$) reflects the product between turnover frequency and the density of attack sites, Γ_{attack} (which will be elaborated below). In other words, $^{inv}V_{max}/S_0$ is a combined measure of the hydrolytic speed and the ability to find attack sites on the cellulose surface, while $^{conv}V_{max}/E_0$ is solely a measure of the maximum rate achievable for each enzyme (Kari et al., 2017).

The parameter Γ_{max} indicates the total number of binding sites on the cellulose surface (given in µmol/g substrate). This parameter can be derived from the traditional Langmüir equation, $\Gamma = \Gamma_{max} \cdot \frac{E_{free}}{K_d + E_{free}}$, and reflects the maximum adsorption capacity, where Γ is the bound enzyme per gram cellulose defined as $\Gamma = \frac{E_0 - E_{free}}{S_0}$. These parameters can be derived by fixing the substrate concentration and varying the enzyme load. By simply measuring the amount of free enzyme present at the different ratios after incubation and plotting bound enzyme ($E_{bound} = E_0 - E_{free}$) as a function of E_{free}, the Langmüir equation can be fitted the results and the parameters derived. Furthermore, the partitioning coefficient, K_p, can be calculated as $K_p = \Gamma_{max}/K_d$, representing the initial slope of the isotherm. However, the assumption underlying the Langmüir isotherm is an undiscriminating measure of binding sites. If we assume that different binding sites are available on the cellulose surface, then this is not the ideal way to interpret this. Nevertheless, the Langmüir isotherm is a widely used method that usually fits data well and provides operational parameters for e.g. comparative analysis (Zhang & Lynd, 2004).

As previously mentioned, cellulose's crystalline nature results in different accessibility of the binding sites. Some attack sites might be easily accessible, whereas some might be completely buried in the crystal hydrogen bonded to all sides and some might be somewhere in between. Γ_{attack} is a parameter that roughly reflects the number of attack sites on the cellulose surface per gram of substrate (µmol/g). We can approximate this number, if we assume the catalytic rate is equal during both substrate and enzyme excess, and we define Γ_{attack}, based on the kinetic parameters, as $\Gamma_{attack} = \frac{^{inv}V_{max}/S_0}{^{conv}V_{max}/E_0}$ (Kari et al., 2017). As mentioned above Γ_{attack} indicates the number of attack sites on the cellulose surface, whereas Γ_{max} indicates both productive

(attack sites) and unproductive binding sites. An unproductive complex could occur from a situation where an enzyme is only associated by the CBM while no substrate is in the tunnel, or it could be an enzyme not able to move forward because of an obstacle (Igarashi et al., 2009; Jalak & Väljamäe, 2010; Praestgaard et al., 2011). A ratio of these parameters, $\Gamma_{attack}/\Gamma_{max}$, reflects how many sites can form a productively competent complex and indicate the efficiency and selectivity of an enzyme.

2.3.1 Assumptions

For both models, we assume the conditions to be in steady-state and use the quasi-steady-state assumption. Prior to this is the pre-steady-state that has previously been observed and described for TrCel7A (Praestgaard et al., 2011), where a fast reaction rate is observed, after which it levels off entering the steady-state phase. To determine a good time for measuring the "initial rate", one needs to weigh the considerations. On one hand, that the reaction has not proceeded too far, thus, converted too much substrate, or produced too much product to avoid product inhibition as well as the slow-down of the steady-state rate, which is a well-known phenomenon for cellulases (Bansal, Hall, Realff, Lee, & Bommarius, 2009). On the other hand, that the state is not too close to the burst phase as well as detectable amounts of product has been produced. This research group has examined this carefully and found, that end-point measurements after one hour is a good compromise of these considerations.

Kinetic parameters are dependent on the specific substrate because the substrate accessibility varies from type to type. As an example, we can use crystalline cellulose versus amorphous cellulose. On the crystal, only a fraction of the ends will be readily available for cellobiohydrolases. Above we introduced Γ_{attack} as a way to define this, but knowing the DP and dimensions of the crystal could give a rough estimate of the attack sites, but probably entail large deviations. Moreover, this would be even harder to estimate for amorphous parts of a substrate. Furthermore, diversity between the amounts of interchain hydrogen bonds (between the cellulose strings) of the specific substrate also contributes to the magnitude of the kinetic parameters. Even on a pure cellulose crystal, the energy needed to pull op a cellulose chain from different places on the crystal varies (Payne, Himmel, Crowley, & Beckham, 2011). Because of cellulose's insolubility and the difficulties of estimating a molar concentration of the attack sites, we assume that attack sites in initial rate measurements scale proportionally with substrate loads, as well as being constant

within a low amount of conversion (Cruys-Bagger et al., 2013). After peeling off a cellulose chain, we imagine that a new chain end will appear, and therefore can the conservation of the attack sites be maintained.

Concerning the concentration of enzyme added to reactions or measured from supernatants, we assume that we can calculate the molar concentration using an equation to quantify the fluorescence signal we measure into number of enzymes (Pace, Vajdos, Fee, Grimsley, & Gray, 1995). We also assume the enzyme to be conserved during the reaction as well unfolding should not be a problem as we stay well below denaturing temperatures. Lastly, we also neglect product inhibition, as the amount of product released during the reaction period is below the inhibition constants reported for the applied enzymes (Murphy et al., 2013; Teugjas & Väljamäe, 2013).

In order to investigate the rate limiting steps, it is interesting to look at cellulase kinetics within short reaction times. In the following section, we will touch upon the rate of threading determined by a newly develop fluorescence based method.

2.4 The rate of cellulose threating determined by real time fluorescence

For catalysis of the TrCel7A to occur, a cellulose strand needs to enter the catalytic binding tunnel and this process is henceforth called 'threading' or 'complexation'. Binding of the CBM to the cellulose surface is therefore not what is referred to using this term, it would rather be 'association'. We developed a method to measure the tunnel being filled with substrate by utilizing the enzymes intrinsic fluorescence (**paper III**).

A library of nine enzymes were cloned and expressed. It entailed the TrCel7A wt and the inactive counter partner E212Q (Ståhlberg et al., 1996) and the catalytic domain without linker and CBM (Core). Further, we mutated the tunnel tryptophans to alanine singularly (W38A, W40A, W367A and W376A) as well as made two double mutants W38A+W376A and 367A+W376A. Using our newly developed method (described below), we were able to measure the speed of the cellulose being threaded as well as the dynamics of the tunnel tryptophans.

2.4.1 Measuring on-rate with tryptophan "de-quenching"

The fluorescing amino acids of a protein, dictates its fluorescence emission spectra. While focusing on tryptophans by exciting only those, it is possible to observe small molecular alterations in the microenvironment as the specific emission spectra changes. In TrCel7A, four tryptophans are lining the binding site. They are directed towards the inner space of the tunnel, and therefore exposed to alterations of the environment upon substrate binding. The emission of an empty binding tunnel will be quenched, as water (a quencher of fluorescence signal) will be filling the tunnel. After threading and binding a cellulose chain, polar water molecules will be excluded from the indole ring of the tunnel tryptophans and the emission signal is de-quenched. Therefore, an increase of the fluorescence signal upon substrate binding will happen, as has been observed previously with the same enzyme by Boer and Koivula (2003).

By using an excitation wavelength of 295 nm we almost exclusively excited tryptophans and not tyrosines or phenylanaline (Lakowicz, 2013). A pre-study was done in order to determine the wavelength of the emission to be used for time cause

measurements. Emission spectra of the wt (empty tunnel) displayed a peak at 342 nm, while the saturated enzyme (fully occupied tunnels) had increased intensity, as a response of de-quenching. Upon binding the substrate, a blue shift towards a lower wavelength was displayed (Figure 13, left). The wavelength 328 nm was selected for time course measurements, as the intensity change was most pronounced here. Time cause measurements (Figure 13, right) were performed on all the enzymes with different RAC concentrations as well as titrations for recording of the threading speed and the saturation level.

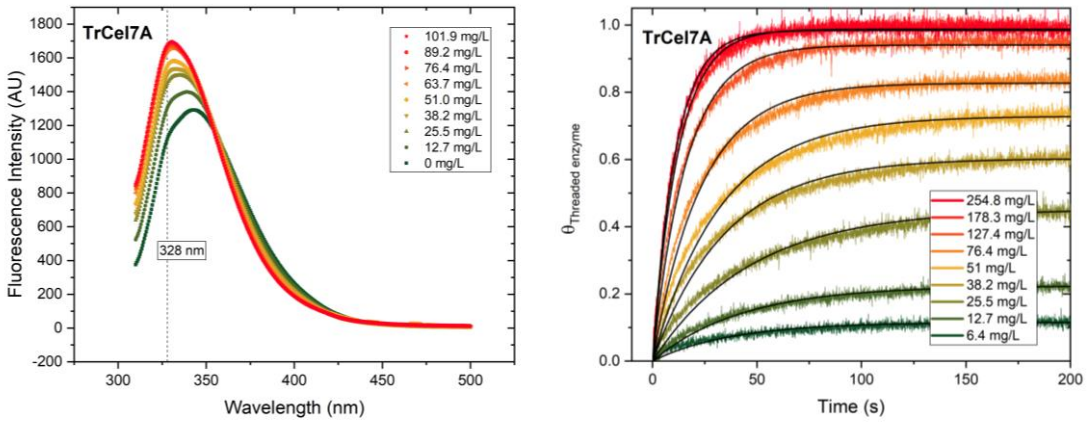


Figure 13: Left: Emission spectra of TrCel7A wt (238 nM enzyme) with no substrate and several different RAC concentrations. Dotted lines are displayed to aid the eye. Right: Time course measurements depicting TrCel7A getting threaded with substrate as the fluorescence signal increases. The data is fitted (black line) with a single exponential function. The y-axis indicate the portion of enzyme on complex with substrate. The substrate was RAC, and the final concentrations appear from the legend. Data from **paper III**.

The initial slope of the time course measurements (within the first few seconds) (Figure 13), represents the rate of the enzymes being threaded per time (nM/s). This is because change in signal emitted represents only complexes where the cellulose string is in the tunnel of the enzyme. We are therefore solely measuring the rate of complexation, v_{on}, for productively bound enzymes, and not for instance enzymes adsorbed only by the CBM, as such association would not give rise to a change in signal, since no tryptophans are present there. As the concentration of RAC added was increased, so was the measured rate of complexation (the initial slope increases, see Figure 13). Plotting the v_{on} as a function of the actual substrate concentration, resulted in a straight line, confirming that the reaction was first order with respect

to the substrate in the measured range. As the on-rate may be written as $v_{on} = {}_pk_{on}S_0E_0$ of which the slope, a, is expressed as $a = {}_pk_oE_0$ and the on-rate constant, ${}_pk_{on}$, can therefore be derived as the slope of Figure 14 when the E_0 is known (all constants are presented in Table 2).

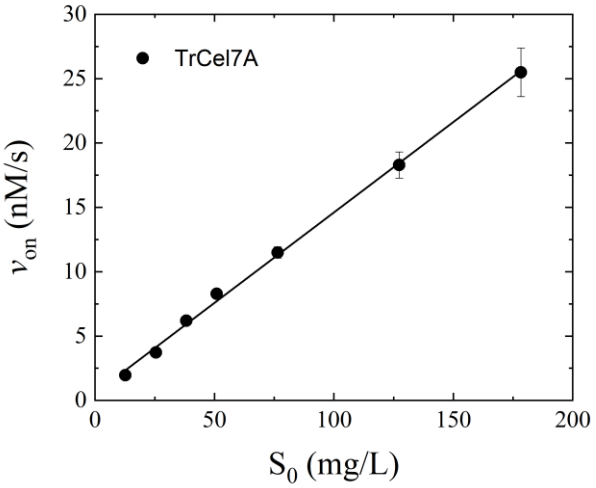


Figure 14. A representation of the on rates (v_{on}) as a function of substrate (RAC) concentration. v_{on} was extracted from Figure 13, right, and plotted against the given RAC concentration. The slope of this line is the on-rate. Figure based on data from **paper III**.

The ${}_pk_{on}$ derived from this plot is in the less common mass unit, $(L/g) \cdot s^{-1}$, as the substrate load is in g/L, and this will henceforth be termed ${}^{mass}{}_pk_{on}$, Table 2. We found that the wt and the catalytically deficient variant E212Q displayed similar ${}^{mass}{}_pk_{on}$, nicely supporting the validity of the method as it is expected that the E212Q does not affect the on-rate constant. More interestingly, we found that ${}^{mass}{}_pk_{on}$ for all the single mutation variants were similar to the wt or moderately decreased revealing that substitution of one tryptophan with alanine does not influence the on-rate on amorphous cellulose significantly. This suggests that the conservation of the tryptophans may not be driven by maintenance of high on-rate. Substitution of two tryptophans to alanine did however hamper the observed ${}^{mass}{}_pk_{on}$ with a halving of the on-rate.

Using the same fluorescence based technique, we were able to make binding curves and find the density of attack sites on RAC, Γ_{attack}, from titration-measurements (see

paper III for further information). These results were very much in accordance with earlier studies (Karuna & Jeoh, 2017; Pellegrini et al., 2014). By this parameter, we could convert the mass units of the on-rate to the more known molar units (${}_pk_{on} = {}^{mass}{}_pk_{on}/\Gamma$), which as a consequence scales with the found number of density of attack sites. From the on-rate constant and K_p we derived the off-rate, ${}_pk_{off} = {}_pk_{on}/K_p$, (Table 2) (**paper III**). In general, these numbers reflect the slow dissociation from the substrate compared to the complexation, and determine the dissociation from the substrate as the rate limiting step at saturation (Cruys-Bagger et al., 2012; Kurašin & Väljamäe, 2011).

Table 2. Rate constant for threading derived from plots like Figure 14, above, and derived on- and off-rates. Data from **paper III**.

Enzyme	**Real-time exp.**	**Derived parameter**	
	${}^{mass}{}_p\boldsymbol{k}_{on}$ **(L/g s^{-1})**	${}_p\boldsymbol{k}_{on}$ **(mM^{-1} s^{-1})**	${}_p\boldsymbol{k}_{off} \cdot \mathbf{10^{-3}}$ **(s^{-1})**
WT	0.59 ± 0.01	93 ± 9	4.8 ± 1.3
E212Q	0.63 ± 0.03	94 ± 11	9.9 ± 2.0
Core	0.33 ± 0.02	163 ± 16	5.9 ± 2.8
W38A	0.41 ± 0.06	157 ± 46	16.6 ± 8.1
W40A	0.39 ± 0.01	94 ± 13	9.4 ± 3.0
W367A	0.59 ± 0.05	85 ± 10	11.1 ± 1.8
W376A	0.43 ± 0.06	48 ± 8	2.2 ± 0.7
W367A+W376A	0.26 ± 0.01	27 ± 4	2.6 ± 0.6
W38A+W376A	0.23 ± 0.02	32 ± 9	3.7 ± 1.7

Summarizing, we developed a fluorescence based real-time method for determination of rate constants. This simple determination of the mass based on-rate is a robust method to directly measure the on-rate. The complexation is a fast process, reaching its half-saturation within about 10 s, and we also found that single mutations of tryptophans do not change the threading rate drastically. In the following chapter, we will use steady-state kinetics for characterization of the properties and functions of several variants with modifications in the loop regions.

3 Exploring the properties of TrCel7A loops

Loops has been ascribed to a lot of specificities and functions (Hobdey et al., 2016; Meinke et al., 1995; Sørensen et al., 2017), such as determining the function between cellobiohydrolases and endoglucanses (Meinke et al., 1995) and directing the efficiency of the enzyme (Sørensen et al., 2017; L. E. Taylor et al., 2018). Those studies has used various methods for testing and elucidating the properties of the peripheral loops. Such methods could be swapping entire modules or loops, amino acid substitution (site saturation) of a specific hot spot as well as during rational engineering on the basis of bioinformatics and getting inspired by other enzymes with known and desirable properties or functions.

Years of research has also taught this group that changes in the peripheral loop regions give raise to quite large changes of the enzymes properties. By employment of various protein engineering strategies, we have investigated the properties of several loops of *Trichoderma reesei* Cel7A, of which some will be presented in this chapter.

3.1 Kinetic differences between Cel7A and Cel7B

The industrial cocktail for depolymerization of lignocellulosic materials contains two major GH7 enzymes, being Cel7A and the Cel7B. In the model fungus *T. reesei,* these show high sequential and structural similarities but functionally they are quite different (Kleywegt et al., 1997; Knott et al., 2013; Penttilä et al., 1986). While Cel7A is the exo-acting processive cellobiohydrolase showing specificity for crystalline cellulose (Cruys-Bagger et al., 2012; Knott et al., 2013; Kurašin & Väljamäe, 2011; Praestgaard et al., 2011; Ståhlberg et al., 1996), Cel7B is the endo-acting endoglucanse with high specificity for the amorphous parts of cellulose (Kleywegt et al., 1997; Zhang & Lynd, 2004).

Having a look at the sequences of these enzymes reveal several gaps in the TrCel7B equal to flexible peripheral loops (Figure 15). The gaps are suggested to arise from deletions from the Cel7A backbone (Kleywegt et al., 1997) and leads the thoughts towards divergent evolution. On DNA level, one could imagine that the Cel7A gene was duplicated followed by several deletion events resulting in two different genes. One encoding the mother enzyme (Cel7A) and one encoding a new enzyme (Cel7B) with functions that improved the organisms fight for survival.

The eight loops in TrCel7A (A1-A4 and B1-B2) are located on each side of the substrate-binding area. The loops missing in the EG corresponds to B2, B3, B4 and A4 (see Figure 15A) and the differences result in two very diverse active sites, namely the cellobiohydrolase's tunnel and the endoglucanase's cleft (Figure 15B).

A

B

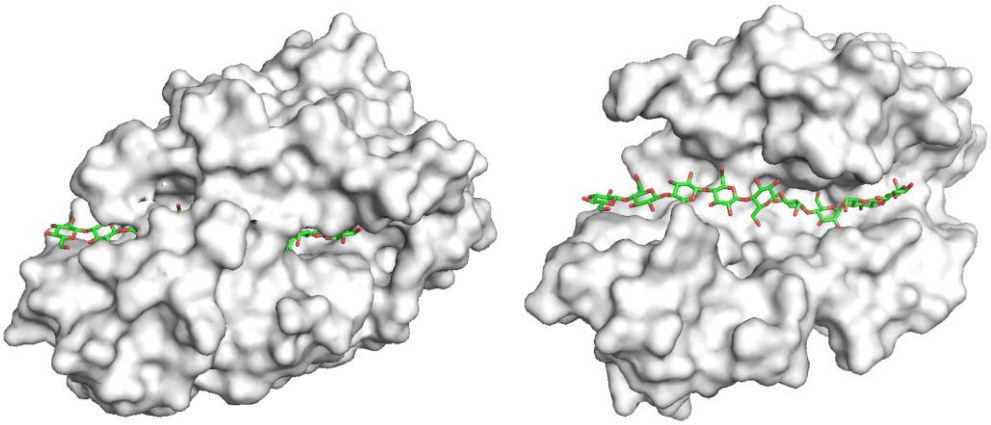


Figure 15: **A)** Pairwise alignment of TrCel7A and TrCel7B amino acid sequences. The loops unique to Cel7A – thus missing in Cel7B, is highlighted in colored boxes. **B)** Surface structure of TrCel7A (left) (PDB 4C4C, (Knott et al., 2013)) and TrCel7B (right) (PDB 1EG1, (Kleywegt et al., 1997) core domains with the cellononaose of 4C4C). The figure illustrates their quite different binding area.

We systematically characterized these two functionally different wild types kinetically, and made a group of intermediate enzymes by swapping loops between them. We did this, to investigate the structure-function relations of the loops that are unique to TrCel7A and TrCel7B. In order to convey the kinetic findings of the variants made in this study, the two "counterparts", the wild types, will be outlined first to create a grid to "hang" the following results at. An introduction of the variants will be provided later.

Conventional MM kinetics (Figure 16), was investigated by saturating a low and constant enzyme concentration with different substrate concentrations. Experimental results for TrCel7A were similar to earlier results (J. P. Olsen et al., 2016), and we found that the turnover rate and the K_M for the Cel7B was ten times higher than for the CBHI. This relation was in agreement with the interpretation of maximum speed being restricted by slow dissociation. In other words, this correlation builds on the theory of Cel7A maximal rate being restricted because of it being stalled in front of an obstacle on the substrate surface (Jalak & Väljamäe, 2010; Praestgaard et al., 2011). Based on that, the strong binding (low K_M) between the cellulose and TrCel7A results in a lower ${}^{conv}V_{max}$ (Kari et al., 2014). Contrary, the Cel7B displayed a higher

$^{conv}V_{max}$ as a consequence of the ten times lower affinity (10 x higher K_M), and thereby a faster dissociation (higher off-rates) (Figure 16). In continuation of our findings, the specificity constant, $^{conv}\eta$, (ratio of $^{conv}V_{max}/(^{conv}K_M \cdot E_0)$), was surprisingly the same for the wild types (Table 3).

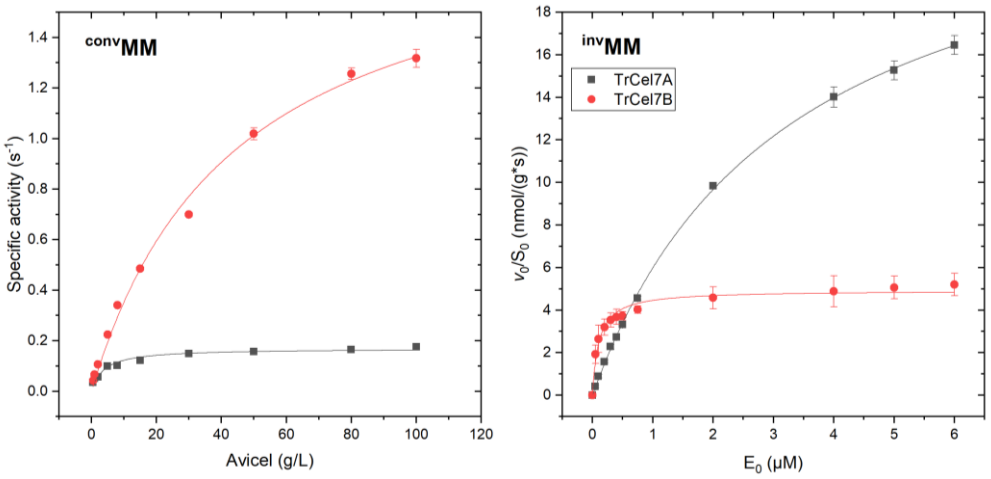


Figure 16: Conventional MM (left) and inverse MM (right) of TrCel7A and TrCel7B. Figure prepared from data from **paper I**.

Table 3. Kinetic parameters of TrCel7A and TrCel7B. Data from **paper I**.

	Conventional MM			Inverse MM		Kinetic substrate accessibility
Enzyme	$^{conv}V_{max}/E_0$	$^{conv}K_M$	$^{conv}\eta$	$^{inv}V_{max}/S_0$	$^{inv}K_M$	Γ_{attack}
	(s^{-1}) $\times 10^{-3}$	(g/L)	($Lg^{-1}s^{-1}$) $\times 10^{-3}$	(µmol $g^{-1}s^{-1}$) $\times 10^{-3}$	(µM)	(µmol/g)
TrCel7A	172 ± 6	4 ± 0.7	43 ± 8	25 ± 0.4	3.2 ± 0.1	0.147
TrCel7B	1908 ± 97	44 ± 5.1	43 ± 5	5 ± 0.1	0.1 ± 0.0	0.003

For the **inverse MM** approach the experimental conditions are switched (Figure 16). Here, the enzyme is in excess while saturating a low, constant concentration of substrate resulting in additional enzyme in solution (Andersen, Kari, Borch, & Westh, 2018; Kari et al., 2017). An enzyme capable of forming catalytically competent complexes on many different sites, will show a high $^{inv}V_{max}$ as it scales with the number of attack sites for the specific enzyme. This kind of enzyme could

be described as "promiscuous" regarding the sites it "selects", in contrast to enzymes that have a lower ability to find catalytically competent sites. The latter can be described as a "selective" enzyme. We can roughly estimate the substrate accessibility, number of attack sites/gram cellulose, by calculating the ratio between the inverse and conventional maximum speed $\Gamma_{\mathbf{attack}} = \frac{^{\mathrm{inv}}V_{\mathrm{max}}/S_0}{conv_{V_{\mathrm{max}}/E_0}}$ (Kari et al., 2017). Γ_{attack} was found to be 0.147 µmol/g for TrCel7A and 0.003 µmol/g for TrCel7B. This estimates the TrCel7A to be 50 times better at finding attack sites (Table 3). Some general properties of TrCel7A and TrCel7B summarized in Table 4 outlining their functional differences.

Table 4. General overview of some properties of TrCel7A and TrCel7B.

Property	Cel7A	Cel7B
Number of attack sites	High	Low
Affinity	High	Low
Processivity	High	Low
Product inhibition	High	Low
Endolytic activity	Low	High
Turnover frequency	Low	High

3.2 Systematic deletions in TrCel7A reveal flexible loop critical for CBH activity

We constructed variants where we singularly deleted the loops corresponding to the ones missing in TrCel7B (B2, B3, B4 and A4) (see Figure 17 and **paper I**). We made a library of variants where we deleted 4 to 14 amino acids. Three of them were different deletions of the B2 loop. As this loop interacts with several other loops, we wanted to make sure that enough stabilizing interactions were maintained in at least one of them in order to obtain a functional enzyme. These were named ΔB2-1 (ΔW192-G205), ΔB2-2 (ΔE192-G205) and ΔB2-3 (S196-T201) and were all successfully made. We also successfully cloned ΔB3 (ΔG245-G253), ΔB4 (ΔE335-F338) and ΔA4 (ΔN384-S388), whereas ΔA4 unfortunately turned out to be kinetically deficient. Moreover, we also made insertions into the Cel7B of some of the

"missing" loops. They were named iA4, iB4, iA4B3 and iB1B2A3, and the work concerning these is unpublished, but the main finding will be presented in further detail later in this chapter.

A

```
TrCel7A   368 DDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVES 409
ΔA4       368 DDYYANMLWLDSTYPT-----TPGAVRGSCSTSSGVPAQVES 404
TrCel7B   321 NDNSQYMNWLDSG------------NAGPCSSTEGNPSNILA 350

TrCel7A   177 PRDLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEA 218
ΔB2-1     177 PRDLKFINGQANVEG--------------HGSCCSEMDIWEA 204
ΔB2-2     177 PRDLKFINGQANVEGW-------------HGSCCSEMDIWEA 205
ΔB2-3     177 PRDLKFINGQANVEGWEPS------GIGGHGSCCSEMDIWEA 212
TrCel7B   176 PVQTWRNGTLNTSHQ---------------GFCCNEMDILEG 203

TrCel7A   226 TPHPCTTVGQEICEGDGCGGTYSDNRYGGTCDPDGCDWNPYR 267
ΔB3       226 TPHPCTTVGQEICEGDGCG---------GTCDPDGCDWNPYR 258
TrCel7B   210 TPHSCTATA---------------------CDSAGCGFNPYG 230

TrCel7A   319 GSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGM 360
ΔB4       319 GSYSGNELNDDYCTAE----GGSSFSDKGGLTQFKKATSGGM 356
TrCel7B   288 DTISSCPSASAYG----------------GLATMGKALSSGM 313
```

B

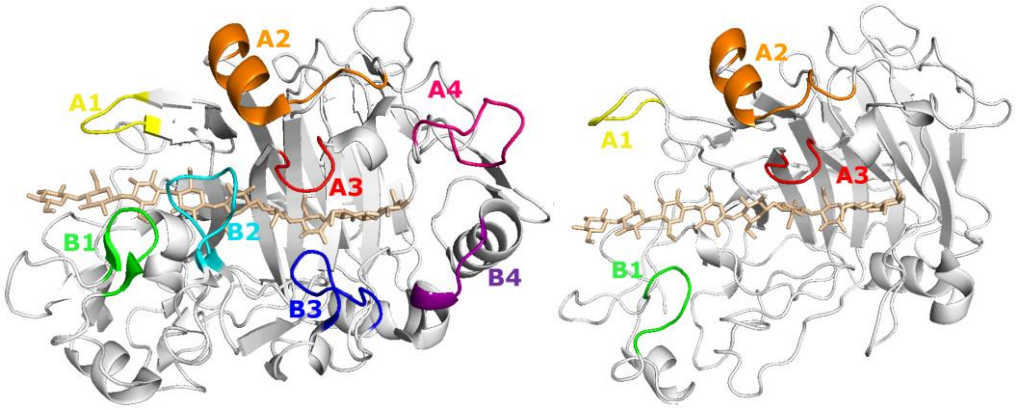


Figure 17: **A)** Amino acid sequence alignment of TrCel7A, the deletion variants and TrCel7B. The colored piece represents the sequence assigned the loop (nomenclature by Momeni et al. (2013)), and the dashed lines represent sequence not present in the variant compared to TrCel7A. **B)** Cartoon of TrCel7A (left) and TrCel7B (right) and their loops (colored). PDB: 4C4C and 1EG1 with the cellononaose from 4C4C aligned. Made with Pymol.

Deleting the **B2 loop**, which covers the -4 and -3 subsites, systematically resulted in kinetic parameters shifting from the CBH towards the characteristics of an EG (Figure 18). Variants with three different degrees of truncations of the B2 loop, showed significantly increased K_M as well as maximum specific rate. The binding

isotherms revealed that binding, Γ_{max}, Table 5, did not change significantly from Cel7A, indicating that the tested loop did not govern the absorption to Avicel in itself. The loop rather seems to govern the ability to form productively competent complexes, as the Γ_{attack} was reduced significantly by tenfold. As well as for the wild types, the variants had almost unchanged specificity constant, implying that the enzyme variants were in good conditions even after the radical changes we made. Furthermore, these findings also suggest that there is a tradeoff between speed and affinity and it seems that changes to the enzyme's specificity constant is not governed by the tested loops. Altogether, these results suggest that the B2 loop is particularly important for TrCel7A to function as a cellobiohydrolase.

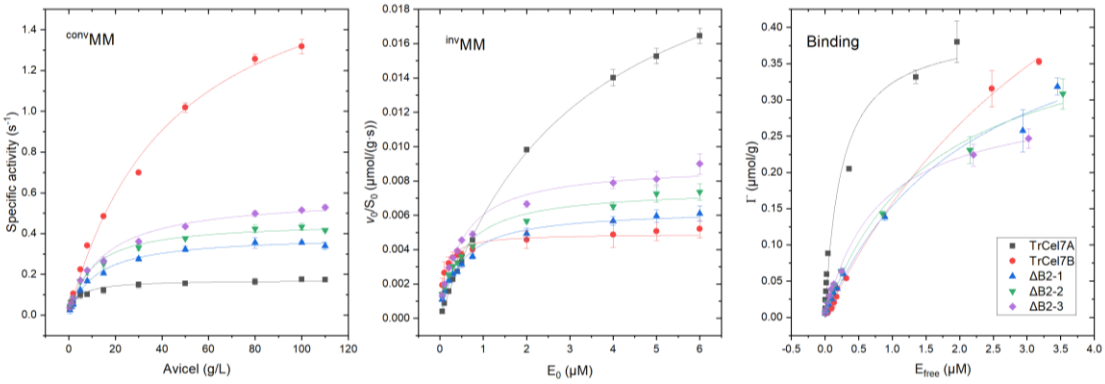


Figure 18: Conventional MM (left) and inverse MM (middle) and binding isotherms (right) of TrCel7A and TrCel7B and the B2 variants. Figure prepared from data from **paper I**. Conditions were as follows: ConvMM (0-110 g/L Avicel, 50 nM enzyme, 60 min 25 °C), InvMM and binding (8 g/L Avicel, 0-6 µM enzyme, 60 min and 25 °C).

Table 5. Parameters of conventional and inverse MM and binding parameters in addition to attack site density of TrCel7A, TrCel7B and the variants ΔB2-1, ΔB2-2, and ΔB2-3 using Avicel as substrate. Data form **paper I**.

Enzyme	convMM		invMM		Kinetic substrate accessibility	Adsorption isotherms	
	$^{conv}V_{max}/E_0$ (s^{-1}) $\times 10^3$	$^{conv}K_M$ (g/L)	$^{inv}V_{max}/S_0$ (µmol g^{-1} s^{-1}) $\times 10^3$	$^{inv}K_M$ (µM)	Γ_{attack} (µmol/g)	Γ_{max} (µmol/g) $\times 10^3$	K_d (µM)
TrCel7A	172 ± 6	4 ± 0.7	25 ± 0.4	3.2 ± 0.10	0.147	0.40 ± 0.03	0.24 ± 0.06
ΔB2-1	391 ± 8	12 ± 1.0	6 ± 0.2	0.5 ± 0.06	0.016	0.48 ± 0.04	2.13 ± 0.41
ΔB2-2	458 ± 12	10 ± 1.0	8 ± 0.3	0.5 ± 0.07	0.016	0.43 ± 0.04	1.56 ± 0.33
ΔB2-3	583 ± 19	15 ± 1.8	9 ± 0.3	0.5 ± 0.06	0.015	0.38 ± 0.03	1.32 ± 0.29
TrCel7B	1908 ± 97	44 ± 5.1	5 ± 0.1	0.1 ± 0.02	0.003	0.87 ± 0.10	4.60 ± 0.78

Deletion of the **B3 and B4 loops** did not change the kinetic parameters significantly from the TrCel7A, thus, indicating that these loops are of less importance for the CBH-like behavior. Von Ossowski et al. (2003) previously constructed a variant removing the B3 loop where they deleted one less amino acid compared to the variant we constructed. They had it crystalized, and found no major structural changes. This again supports the impression of our enzymes being properly folded. Deletion of five amino acids in the **A4 loop**, turned out to be destructive and resulted in a decrease of the T_m (DSC measurements) by 10 °C as well as drastic lowering of the activity.

3.2.1 Specificity to other cellulose substrates

The aforementioned kinetic studies were performed using Avicel as substrate. This was a fair choice, as Avicel is approximately 50 % amorphous and 50% crystalline cellulose (Zhang & Lynd, 2004). As mentioned previously, the Cel7A prefers the crystalline part and the Cel7B prefers the amorphous part (Jalak, Kurašhin, Teugjas, & Väljamäe, 2012; Ståhlberg et al., 1993; Zhang & Lynd, 2004) and therefore, Avicel seems to be a good model substrate for both enzymes. However, the activities of the variants were tested on two other cellulose substrates; regenerated amorphous cellulose (RAC) (Zhang et al., 2006), which is primarily amorphous cellulose and bacterial microcrystalline cellulose (BMCC), which as the name indicates is predominantly crystalline cellulose with a DP around 100 (Väljamäe et al., 1999).

Our results reflected the literature concerning the preferences mentioned above, and saw that deletion of the B2 loop resulted in enhanced performance on the EG-preferred-substrate, RAC, indicating that also on this parameter the deletion shifted its property from the CBHI to the EG. Additionally, we also observed loss in activity on the CBHI-preferred-substrate, BMCC, supporting this shift (Figure 19).

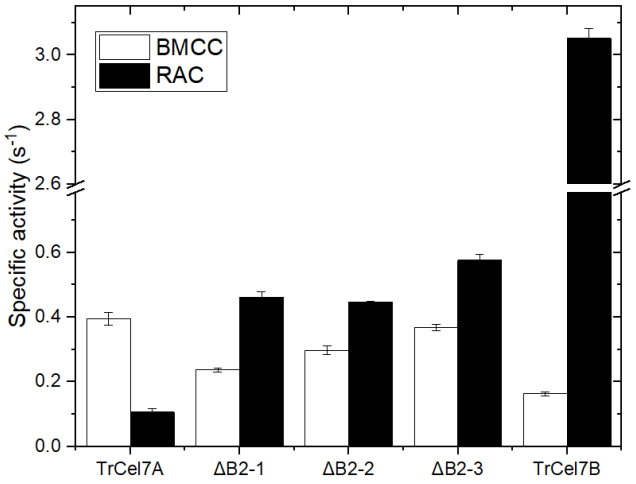


Figure 19: Bar plot visualizing specific activity (v/E_0) for end-point conversion of BMCC (white) and RAC (black). Conditions: 4 g/L RAC or BMCC and 50 or 100 nM enzyme, respectively, at 25 °C for 60 min. Figure based on data from **paper I**.

3.3 Insertion of loops into the Cel7B

In order to strengthen and further support the hypothesis of the loops governing the properties of the cellobiohydrolase, we wanted to show that insertion of loops into the endoglucanase likewise would introduce the properties from the CBH to the EG. However, inserting loops consisting of many amino acids into the backbone of the Cel7B was a challenge, as many of them are stabilized by interactions via their neighboring loops. For that reason, we had to swap some existing loops with the one from Cel7A in order to establish sufficient stabilization, as single loop insertion seemed difficult for some constructs. With this method, we moved the stabilizing interactions along, to increase the chance of establishing the right interactions. In some cases, we also only mutated specific amino acid, making the necessary interactions possible. Insufficient stabilization of such loops could result in fluctuations, targets for proteolytic attack or misfolding and these reasons might also be the cause of some of the lost activities we observed in the results.

All enzymes with insertions of loops into the backbone of the Cel7B were named "i" as a prefix followed by the name of the loop. For example, iA4B3 (insertion of A4 and B3 loop from TrCel7A into TrCel7B). We inserted 12 amino acids in iA4, 5 in iB4, 21 plus 12 in iA4B3 and 15 in iB1B2A3 in addition to other mutations as shown

on Table 6. Concerning the variants where the A4 loop were inserted, a markedly higher band on the SDS-PAGE gel (not shown) were observed, indicating insertion of the glycosylation site N384 (TrCel7A numbering) as well as actual glycosylation. The four insertion variants are presented in detail in Table 6 below.

Table 6: An overview of the insertion variants. *) indicates insertion of additional amino acids at the specific position. Here, TrCel7B numbering are used. The inserted loop is marked in parenthesis with bold.

Enzyme	Sequence	Comments
iA4	G230R G333T N334Y A335P *335TNETSSTPGAVR (**A4**)	Glycosylation site inserted
iB1B2A3	T50Y V51D N52G G53N G54T V55W (**B1**). *191EPSSNNANTGIGGHG (**B2**). N321D N323Y S324Y Q325A Y326N (**A3**).	B1 and A3 inserted in addition.
iA4B3	A216T T217V A218G *218QEICEGDGCGGTYSDNRYGGT G230R (**B3**). G333T N334Y A335P *335TNETSSTPGAVR (**A4**).	A4 inserted in addition. Glycosylation site inserted
iB4	A296F *296GGSSF (**B4**)	Only B4 inserted

Now that we found the B2 loop to be responsible for many of the cellobiohydrolase properties, it is obvious to think that inserting this loop into the backbone of the Cel7B will result in the properties of Cel7B shifting towards the Cel7A. This could be shown as a higher affinity in return for a decreased turnover frequency as well as moving towards a higher substrate accessibility. In addition to the results shown above, **paper I**, we would likewise not expect any large changes after introducing the B3 and B4 loop.

Analyzing the results, we found that introduction of the B2 loop (along with the B1 and A3 loop, named "iB1B2A3"), into the TrCel7B backbone, actually did move the behavior from EGI towards CBH. Specifically, we found a twofold lowering of K_M along with a threefold lowering of the turnover frequency, (Figure 20 and Table 7) and as the TrCel7B already had a B1 and A3 loop we mainly ascribed this to the addition of the B2 loop. We know from experience that it is easy to destroy an enzyme by lowering both the turnover frequency and the affinity. However, to increase the affinity, here in terms of a lowered K_M, we consider to be an improved property that usually goes hand in hand with lowering of the maximal rate, because of the restricted ability to escape the substrate when an obstacle is encountered.

Insertion of the B4 did not result in noteworthy changes in the MM kinetics, while inserting A4 in both cases showed to lower both affinity (higher K_M) and $^{conv}V_{max}$.

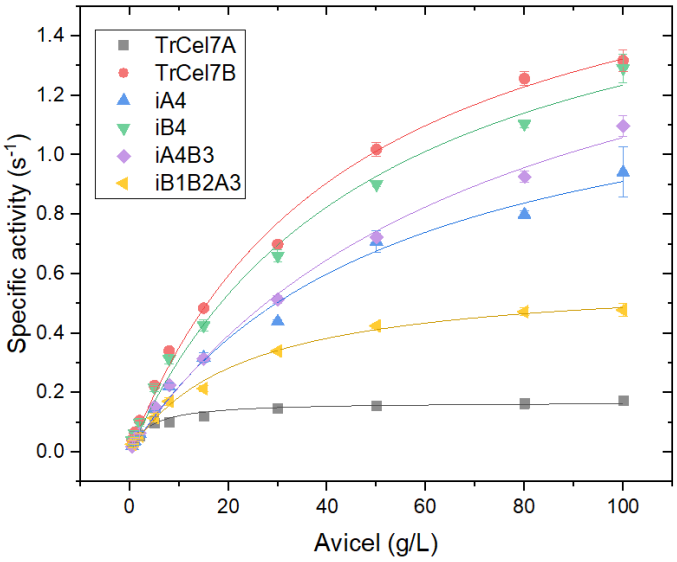


Figure 20: Michaelis-Menten curves of enzymes with inserted loops. The experiments were performed as described in **paper 1**, Conditions: 0-100 g/L Avicel, 50 nM enzyme, 60 min and 25 °C.

Table 7: Kinetic parameters extracted from the Michaelis-Menten plot shown above.

Enzyme	$^{conv}V_{max}/E_0$ (s^{-1})	$^{conv}K_M$ (g/L)
TrCel7B	1.91 ± 0.10	44.15 ± 5.11
iA4	1.39 ± 0.12	52.82 ± 9.27
iB4	1.85 ± 0.12	49.45 ± 7.22
iA4B3	1.84 ± 0.14	73.52 ± 10.57
iB1B2A3	0.60 ± 0.02	22.08 ± 2.12
TrCel7A	0.17 ± 0.01	4.0 ± 0.7

Briefly, the findings from inserting loops into the TrCel7B thought us that insertion of large pieces of amino acids are difficult, and often results in lowering of the stability (T_m). Furthermore, we were able to insert a glycosylation site motif and detect the successful glycosylation by SDS-PAGE. More interestingly, the kinetic results supported our earlier findings from deletion variants. By inserting the B2 loop we

lowered K_M and the specific activity as we expected as well as we found that the B3 and B4 loop were of less importance for the properties. Altogether, swapping the loops between Cel7A and Cel7B strongly indicates the B2 loop to be a key determinant for the properties of these two important cellulases.

3.4 Site saturation in the B2 loop

Based on the previous study, we know that changes in the B2 loop can influence the kinetics of the entire enzyme significantly. We will now zoom in on this kinetically important loop, where we did a site saturation study on N200 in TrCel7A (**Paper IV**) replacing a single position with all amino acids. In order to investigate the function of this much conserved asparagine (Figure 21) and the substitutions, we measured the rate by ^{conv}MM and ^{inv}MM as well as the affinity to Avicel.

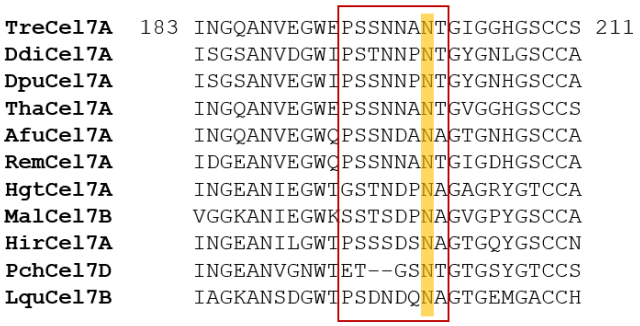


Figure 21: Section of GH7 sequence alignment adapted from (Hobdey et al., 2016). The alignment show the much conserved N200 highlighted in yellow and the B2 loop (red box).

We found that deletion of N200 or substitution with small amino acids (glycine or alanine) resulted in a large increase of the $^{conv}V_{max}$ and likewise also an increase in K_M (Figure 22). This was completely in line with the findings we did after removal of different amounts of the B2 loop (ΔB2-1, ΔB2-2 and ΔB2-3), and these parameters likewise point in the direction of those characterizing an endoglucanase. We also observed that the $^{inv}V_{max}$ of the variants seemed impaired compared to the wt, which was entirely in line with our findings from **paper I**. All the variants showed lower K_p, indicating lower binding abilities, as well as an inferior $^{inv}V_{max}$ (the remaining results can be found in **paper IV**).

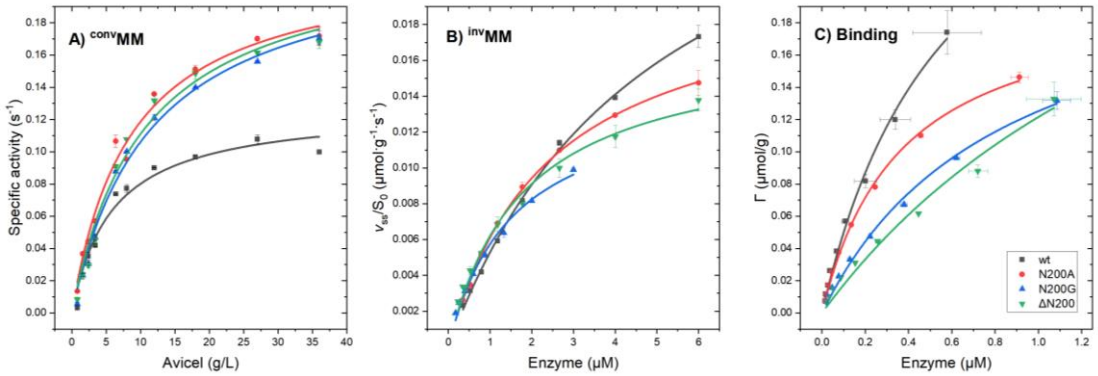


Figure 22: **A)** Conventional MM, **B)** Inverse MM, **C)** Langmüir binding isotherms. Variants showed here: WT, ΔN200, N200G and N200A. Conditions: convMM (38 g/L Avicel, 400 nM enzyme, 25 °C, 60 min) InvMM (13 g/L Avicel, 0-6 μM enzyme, 25 °C, 60 min), Binding (0-1.6 μM, 27 g/L Avicel, 25 °C, 60 min). Data from **paper IV**.

Table 8: Parameters extracted from graphs in Figure 22 above.

	^{conv}MM		^{inv}MM		Kinetic substrate accessibility	Adsorption isotherms
Enzyme	$^{conv}\mathbf{V_{max}/E_0}$ (s^{-1})	$^{conv}\mathbf{K_M}$ (g/L)	$^{inv}\mathbf{V_{max}/S_0}$ (μmol g^{-1} s^{-1})	$^{inv}\mathbf{K_M}$ (μM)	$\mathbf{\Gamma_{attack}}$ (μmol/g)	$\mathbf{K_p}$ (L/g)
TrCel7A	0.13 ± 0.01	5.9 ± 1.1	0.031 ± 0.001	4.9 ± 0.4	0.25	0.58
ΔN200	0.23 ± 0.02	10.4 ± 1.8	0.017 ± 0.001	1.7 ± 0.2	0.07	0.18
N200A	0.22 ± 0.01	8.6 ± 1.1	0.021 ± 0.000	2.4 ± 0.1	0.09	0.56
N200G	0.23 ± 0.01	11.4 ± 1.4	0.014 ± 0.001	1.4 ± 0.2	0.06	0.27

We found that eleven out of the twenty enzymes showed improved $^{conv}V_{max}$ compared to the wild type. Replacement of the N200 with positively charged amino acids were found to increase the $^{conv}V_{max}$, whereas negatively charged amino acids lowered this parameter, while they all resulted in an increased K_M. However, none of the enzymes performed better in the inverse MM. Overall, this means, that many amino acid substitutions lead to a higher speed during high substrate/enzyme ratio (conventional MM), while none of them acquired better properties for finding attack sites or to bind the cellulose in general. From an evolutionary point of view, this indicate that the enzyme has not evolved to make an enzyme with the highest $^{conv}V_{max}$ as this was easy to achieve by substitutions. It seems like evolution rather aimed to make an enzyme that has the best abilities to attack and form productively competent complexes with the cellulose, as we were not able to improve this parameter. This

interpretation is also in agreement with our findings from studying systematic deletions in TrCel7A (**paper I**).

Another study was made investigating a different area of the core domain. This section is kept confidential due to potential patenting, and is therefore remove from here, and will only be available for the evaluating committee. The committee is referred to the confidential appendix.

Summarizing the findings of these systematic studies with focus on the loops and their functions and properties, we found that the B2 loop is of major importance for the CBH-like behavior and removal of large "B2-pieces" drives the kinetics towards those of an EG. Likewise, insertion of the B2 into the Cel7B backbone also suggested that this loop directs the function of the enzyme towards CBH behavior, whereas the B3 and B4 were of minor influence. Additionally, small changes of the B2 loop, such as deletion of N200 or substitution with glycine or alanine, again turns the kinetic parameters towards a more EG-like behavior. Moreover, we found the B2 loop and the asparagine at position 200 to be evolutionary optimized for the TrCel7A to make attack on the cellulose surface.

4 pH effects on cellulases

Traditional enzyme characterization often revolves around pH and temperature optimum, stabilization of the enzyme and activity measurements. Usually, these measurements are performed by determining the optimal temperature followed by the optimal pH or vice versa. Separate determination of these optima can unfortunately exclude important information about the enzymes real pH and temperature profiles (Herlet et al., 2017). We did a comprehensive three dimensional study investigating the relationship between these parameters of some of the most industrially relevant cellulases. In industry, the biomass is typically pretreated with acid or base prior the enzymatic hydrolysis (Jørgensen et al., 2007; Mosier et al., 2005). In addition, the process temperature, as well as the substrate concentration also varies strongly during the industrial hydrolysis, why these dimensions are important to elucidate and consider.

In cellulase studies, small soluble substrate analogs has widely been used as substitutes for insoluble substrates when investigating the pH sensitivity, because of their convenient solubility. These has shown typical bell shaped curves with optimum around pH 4 - 5 (Becker et al., 2001; Boer & Koivula, 2003; Borisova et al., 2018; Hobdey et al., 2016; Momeni et al., 2013; Schülein, 1997; L. E. Taylor et al., 2018), but optima at higher and lower pH have also been reported of both wild types and engineered enzymes (Cockburn & Clarke, 2011; Pellegrini et al., 2015; Texier, Dumon, Neugnot-Roux, Maestracci, & O'Donohue, 2012). Thorough elucidation of the pH optimum on insoluble substrates are more infrequent, but has been done (Borisova et al., 2015; Colussi et al., 2012; Marjamaa, Toth, Bromann, Szakacs, & Kruus, 2013; Ogunmolu et al., 2017).

As performance of an enzyme at specific pH is pivotal knowledge to gain the greatest profit, pH activity profiling are usually part of standard characterization. Table 11 condense a collection of literature that contains pH profile characterization of cellulases and some characteristics of them.

Table 11. A selection of studies including pH activity profiles of fungal cellulases including comments on the pH profile.

Enzyme	pH optimum °C	Substrate	Comments	Reference
Insoluble substrate				
ThCel7A	5	10 g/L Avicel	clearly defined optimum - activity drops drastically on the basic limb	Colussi et al. (2012)
PfCBHI	3.7	10 g/L Avicel	pH-pofile made at 40, 50, 60 and 70 °C. Bell shaped.	Ogunmolu et al. (2017)
GcaCel7A	5	5 g/L Avicel	High activity maintained (>80%) from pH 4-6.5. But bell shaped.	Borisova et al. (2015)
ThCel7B	3	10 g/L CMC	Bell shaped/sharp maximum	Pellegrini et al. (2015)
HiEGI	5.5	7.5 g/L CMC	>80% activity maintained between pH 5.5-7.5	Schülein (1997)
PpCBHI	5	10 g/L Avicel	Bell shaped	Marjamaa et al. (2013)
Soluble substrate				
TrCel7A $TrCel7A_{mut}$	4.5	500 µM 4-methylumbelliferyl-β-D-lactoside (MULac)	clearly defined optimum with apex on both sides	Boer and Koivula (2003)
$DdiCel7A_{CBM}$ $DpuCel7A_{CBM}$ TrCel7A	5 5 4	2 mM pNPL	Bell shaped curves. Cel7A from amoeba. TrCel7A Family 1 CBM added for comparison	Hobdey et al. (2016)
TrCel6A	4 - 6	200 µM cellotetraose	Broad optimum that fluctuate a little. Perhaps because of several buffer systems	Koivula et al. (2002)
PfCel7A + TrCel7A	4	1.6 mM pNPL	Bell shaped	L. E. Taylor et al. (2018)
TreCel7A_CD ThaCel7A_CD TatCel7A_CD	4 4.5 4	2 mM pNPL	Bell shaped. Core domains - prepared with papain	Borisova et al. (2018)
HiCBHI HICBHII	5 9	100 µM cellotriose 100 µM reduced cellohexaose	Bell shaped, six buffer systems used. Bell shaped, but >50 % activity maintained from pH 4-10	Schülein (1997)
HirCel7A	4	1 mM pNPL	Bell shaped	Momeni et al. (2013)
PfCBHI	3.5	10 g/L medium viscosity barley β-glucan	Bell shaped	Texier et al. (2012)

4.1 pH and temperature relations

To begin with, we tested the activity of TrCel7A on high Avicel and on low *para*-Nitrophenyl β-D-lactopyranoside (pNPL) load compared to the respective K_M, and found a striking difference (Figure 26). Our results on pNPL hydrolysis parallels the literature (Borisova et al., 2018; Hobdey et al., 2016; L. E. Taylor et al., 2018) with a text-book example of a bell shaped pH-curve with apex around pH 4.5. Hydrolysis of Avicel (90 g/L), far above the K_M (14 g/L at 40 °C) (Sørensen et al., 2015), surprisingly showed identical activity from pH 2 to 5.5, resulting in a flat horizontal pH profile, with an activity drop from pH 6 and above. From these findings it is reasonable to say that one should be cautious using a substitute as pNPL if one wants to explain the relation between activity and pH on real cellulose.

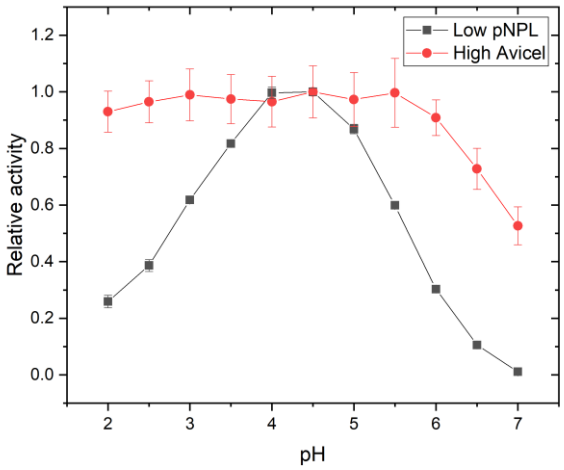


Figure 26. Relative activity of TrCel7A acting on 0.2 mM soluble pNPL and 90 g/L insoluble Avicel at 40 °C as a function of pH. Data from **paper II**.

Next, we investigated the pH-profiles at several temperatures (20 – 70 °C) as well as the T_m over a large range of different pH. The independence of TrCel7A towards changes in pH from 2 to 5.5 was evident when working at 20 °C to 40 °C, see Figure 27A. Stability measurements showed that the enzyme was fully stable at these temperatures (Figure 27 panel C). This is seen from the T_m (red squares) that is approximately 50 °C and higher at all conditions, and we assume that the enzyme is stable at lower temperatures. From the same panel it appears how the enzyme is less stable at the high and low end of the tested pH range. This explains how activity measured at 50 °C and upwards show lower activity in the outermost pH regions, simply because of the thermal instability of the enzyme. In other words, this means that the enzyme is not dependent on the pH from 2 - 5.5 during stabile conditions.

Moreover, this means that the curvature that arise from 50 °C to 70 °C is due to the general instability of the protein at the given pH value, and not slower catalysis *per se*. Contrary, the TrCel7B is at all temperatures dependent on the surrounding pH, even during 20 °C and 30 °C (Figure 27 panel B), well below temperatures that destabilize the enzyme (panel D). TrCel7B cannot withstanding the high temperatures (60 °C and 70 °C) during the one hour hydrolysis, as depicted by the reduced specific activity compared to 50 °C, for instance. This is even though the T_m values are well above 60 °C at pH 3-5, and therefore the enzyme in theory should be stable. However, stability data based on methods where the enzyme is exposed to a given temperature for a short time period (ex. DSF or DSC), can only be used as a hint to the stability of the enzyme during a reaction over longer time periods. Thermal instability during an hour (the specific assay conditions here) is presumably the reason why the activity at 60 °C pH 4.5 is lower than the one on 50 °C, and might be an effect of the open structure of the Cel7B.

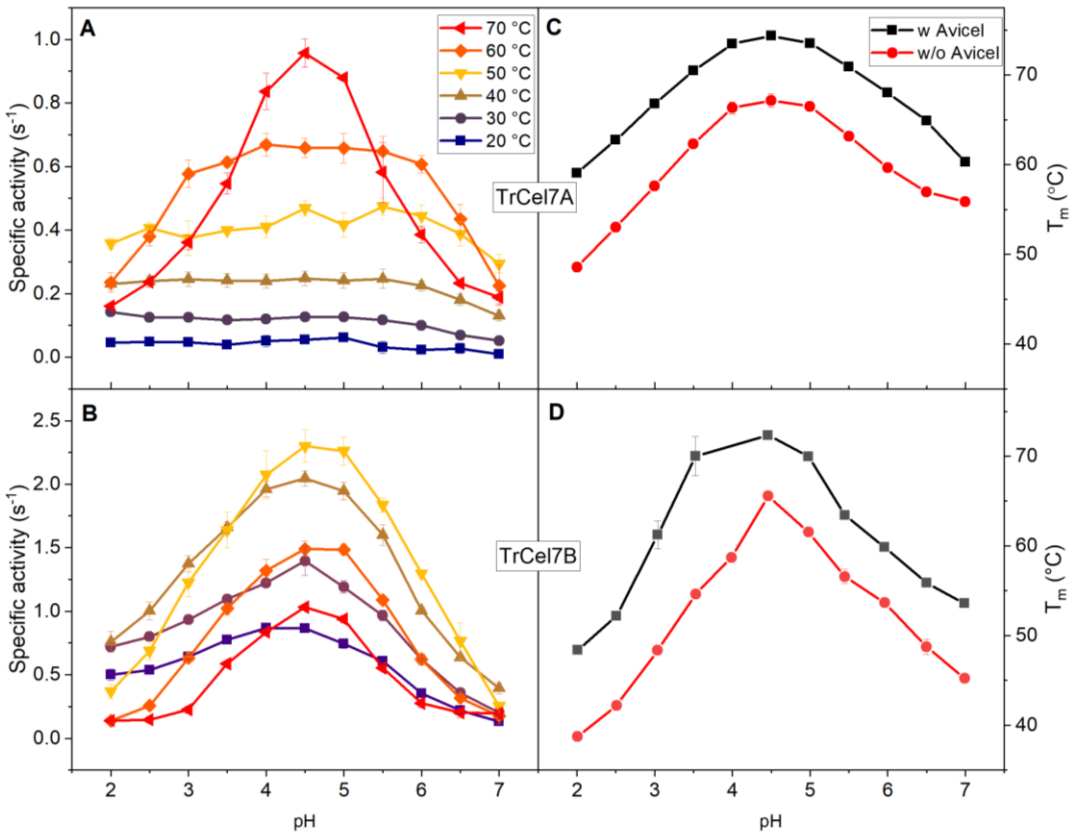


Figure 27. Specific activity for endpoint measurements at different temperatures (20 – 70 °C) and T_m with and without Avicel as function of pH. Results for TrCel7A (A and C) and TrCel7B (B and D). Conditions: 90 g/L Avicel, 100 nM TrCel7A or 50 nM TrCel7B (activity) and 1 µM (T_m) enzyme, 60 min. Data from **paper II.**

4.2 Stabilization increase during binding of cellulose

At all the tested pHs, we found that the presence of 90 g/L Avicel resulted in increased stability of the enzymes (Figure 27 Panel C and D). This improvement is most likely due to the stabilizing effects from the Michaelis complex that forms when the enzyme is bound to substrate and this is a completely ordinary effect as underlined by (Petrauskas, Zubrienė, Todd, & Matulis, 2019). We tested this effect on Cel7A, Cel7B and Cel6A from *T. reesei* and Cel7A from *R. emersonii,* and found that the enzymes gained stability at all tested pH values (Table 12). The two cellobiohydrolases of *T. reesei* were very similar with an improvement of approximately 7 °C from 67 °C to 74 °C without and with Avicel, respectively. The endoglucanase showed a similar increase but with a lower point of origin. The thermophile *R. emersonii* on the other hand, had an almost 10 °C higher T_m without substrate present, but gained only 3 °C in the presence of substrate. We note, that higher T_m will result in a lower amount of enzyme being bound to substrate, and thereby result in a lower increase of T_m (ΔT_m). For instance; two enzymes with similar K_M but with a low or high T_m, respectively, will bind different amounts of substrate, since the amount of bound substrate decrease as temperature increases (similar substrate load for all enzymes). Also, a high K_M of an enzyme will result in lower portion of enzyme being bound to substrate at a given temperature as the affinity to the substrate is lower and less enzyme will be stabilized. Ideally, all enzymes should be bound to substrate when the T_m of the complex is reached, but experimentally, this is not possible as it would require large amount of dry matter that cannot enter the capillary tubes used for nanoDSF.

Table 12: Thermo-stability data of TrCel7A, TrCel7B, TrCel6A and ReCel7A. T_m of TrCel7A, TrCel7B, TrCel6A and ReCel7A where the enzymes are most thermo-stabile, pH 4.5. Average ΔT_m is the average increase of degree Celsius, which complexation with Avicel promotes from pH 2-7. T_m is measured by nanoDSF at pH between pH 2-7 with and without 90 g/L Avicel present. Scan rate 2.0 °C/min, 20 °C - 90 °C, 100 % excitation power, 1 µM enzyme, all measurements performed as duplicates. Numbers are extracted from **paper II**.

Enzyme	T_m w. Avicel (pH 4.5)	T_m w/o Avicel (pH 4.5)	Average ΔT_m
		°C	
TrCel7A	74.4	67.2	7.9
TrCel7B	72.3	65.5	9.2
TrCel6A	74.3	67.6	7.6
ReCel7A	78.6	75.8	3.7

4.3 pH dependency of cellulases with different binding sites

The previously mentioned differences between the loops of TrCel7A and TrCel7B results in an open substrate binding cleft (Cel7B) where water can easily access in contrast to the tunnel of Cel7A, which is less accessible (Figure 28, left) (Bu et al., 2012). In Figure 28 (right), pH activity profiles performed at 40 °C (well below stabile temperatures) has been extracted from Figure 27 and presented relative to the activity at pH 4.5 for TrCel7A and TrCel7B, respectively. The TrCel7B show high dependence on the surrounding pH, whereas TrCel7A is unaffected to whether the pH is 2 or 5.5. Relating this observation to the two enzymes' main structural differences (Figure 28), it is natural to reason that the tolerance towards low pH is related to the accessibility of water in the binding area.

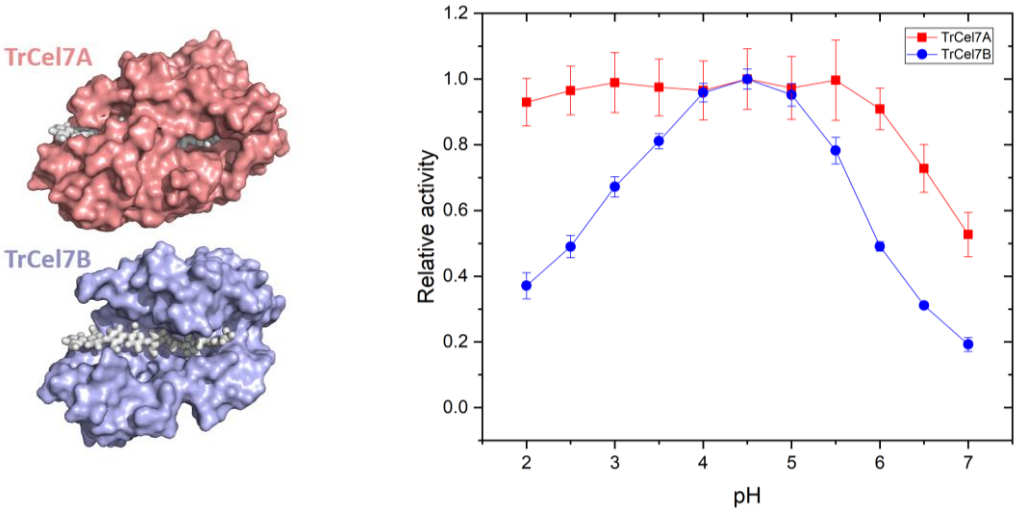


Figure 28. The surface area (left) and pH activity profile at 40 °C on high loads of Avicel (90 g/L) (right) of TrCel7A, 100 nM (red) and TrCel7B, 0.5 nM (blue). Data from **paper II**.

In order for the hydrolysis of these enzymes to be efficient, two acid residues needs to be properly protonated and deprotonated. Changes in surrounding pH can displace these and reduce the efficacy of the catalysis drastically (Harris & Turner, 2002). The molecular understanding of how the TrCel7A can withstand the extremely low pH remains unsolved, but it could be associated to the accessibility of water when the ligand is bound or the number of bonds formed during the Michaelis complexation. The latter has shown to scale with the extent of coverage of the substrate binding

area, so that Cel7A has little water accessibility but forms many H-bonds per subsite (C. B. Taylor et al., 2013).

The amount of interactions in the substrate binding site also seems to have an influence when it comes to the effects of substrate loads. In **paper II** we also investigated different concentrations of soluble and insoluble substrates (pNPL and Avicel). For both substrates, we found that high substrate loads enhanced activity towards lower pH values as well as the cellulose made the enzyme more tolerant compared to pNPL. Hence, during conditions where the substrate concentration was much higher than the K_M – a situation where the binding tunnel of all enzymes are expected to be in complex with substrate - we found a broader optimum plateau as a result of higher tolerance towards lower pH (Figure 29). These findings also emphasized that a large substrate, such as cellulose, in the tunnel results in more tolerance and higher maintenance of activity at low pH compared to the small pNPL analog. We speculate, that when the tunnel is filled by substrate, which span all the subsites, the catalytic residues are protected from the surroundings and therefore we see unchanged activity in a larger pH range.

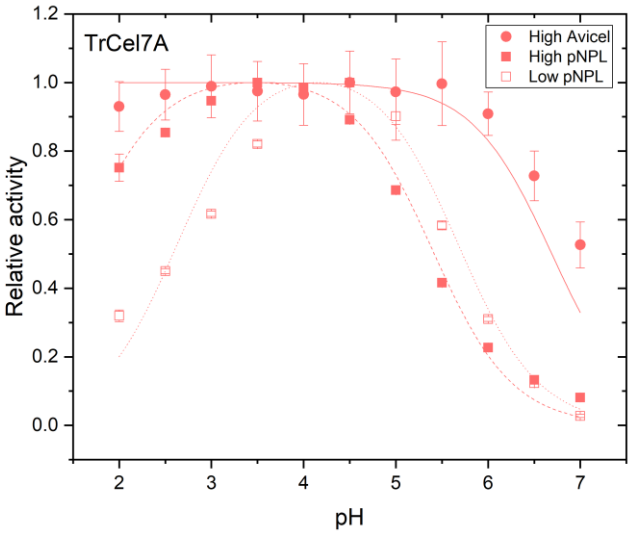


Figure 29: Relative activity of TrCel7A on different substrate loads; high Avicel (90 g/L, 100 nM enzyme), high pNPL and low pNPL (9 mM and 0.2 mM, 500 nM enzyme). Data from **paper II**.

4.4 Michaelis-Menten at different pH

The following data are unpublished and not a part of a manuscript. Therefore, the method will briefly be given here: The measurements were done at 40 °C using the same buffer system as in **paper II** and hydrolysis of Avicel were performed during one hour with concentrations of Avicel from 0 to approximately 100 g/L and 100 nM enzyme. Hydrolysis of pNPL were performed using 100 nM enzyme and pNPL concentrations from 0 to 7 mM for 15 min. Quantification of the cellobiose and pNP concentration were done as in **paper II**.

This conventional Michaelis-Menten characterization of TrCel7A at different pH, substantiate the previous findings of the difference between activity on cellulose and pNPL. A quick glance at the two graphs in Figure 30 reveals how the activity and K_M are similar between pH 2-6 on Avicel, while they are very dissimilar during hydrolysis of pNPL in this range. Further investigations of the parameters from Avicel hydrolysis (Table 13) confirm a drop in activity in the high pH range while the specific maximum rate and K_M is unchanged from 2-5.5 supporting the previously observed plateau. Looking in to the parameters from pNPL hydrolysis, we find that both affinity and rate changes quite a lot. From pH 2 affinity and rate increases until around pH 4, after which they decrease again. Once again confirming the previous observation of the bell shaped pH-profile. These more detailed pH activity investigations surprisingly reveal that kinetically, there is no change over the large pH span (pH 2-6) while hydrolyzing Avicel. In other words, the V_{max} and K_M are practically the same despite the pH of the buffer. Conversely, the specific activities (V_{max}/E_0) on pNPL changes quite a lot, indicating that k_{cat} is influenced by the pH.

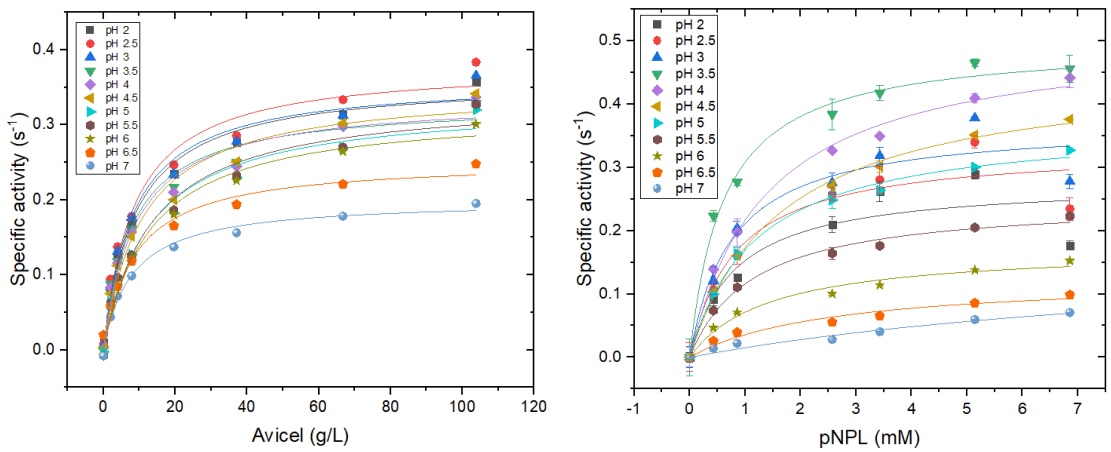


Figure 30: Michaelis-Menten curves of TrCel7A at different pH on pNPL (left) and Avicel (right). Both experiments were performed at 40 °C. Unpublished data, method described in the text.

Table 13. Kinetic parameters from MM kinetics of TrCel7A hydrolyzing Avicel and pNPL. Data extracted from best fit of the MM equation.

	Avicel		pNPL	
pH	K_M (g/L)	V_{max}/E_0 (s^{-1})	K_M (mM)	V_{max}/E_0 (s^{-1})
2	9.52 ± 1.32	0.36 ± 0.01	0.82 ± 0.53	0.28 ± 0.04
2.5	9.09 ± 1.67	0.38 ± 0.02	0.83 ± 0.39	0.33 ± 0.04
3	9.03 ± 1.47	0.37 ± 0.02	0.75 ± 0.31	0.37 ± 0.04
3.5	7.77 ± 1.37	0.33 ± 0.01	0.61 ± 0.08	0.50 ± 0.01
4	9.08 ± 1.85	0.34 ± 0.02	1.37 ± 0.17	0.51 ± 0.02
4.5	10.98 ± 2.21	0.35 ± 0.02	1.64 ± 0.14	0.46 ± 0.01
5	11.55 ± 2.43	0.33 ± 0.02	1.18 ± 0.12	0.37 ± 0.01
5.5	11.68 ± 2.59	0.34 ± 0.02	1.14 ± 0.17	0.25 ± 0.01
6	11.36 ± 1.76	0.32 ± 0.01	1.43 ± 0.34	0.17 ± 0.01
6.5	8.23 ± 1.10	0.26 ± 0.01	2.56 ± 0.94	0.13 ± 0.02
7	6.92 ± 0.73	0.21 ± 0.01	11.07 ± 8.27	0.18 ± 0.09

4.5 Engineering pH stability

Several studies has revolved around shifting pH optimum of cellulases to lower and higher values (Becker et al., 2001; Cockburn & Clarke, 2011; Heinzelman et al., 2009; Wang et al., 2005; Wohlfahrt, Pellikka, Boer, Teeri, & Koivula, 2003). Wohlfahrt et al. (2003) succeeded to increase activities at neutral and alkaline pH by substituting carboxyl-carboxylate pairs with amide-carboxylate pairs in the surface of the loops covering the binding site of TrCel6A. The molecular mechanism behind this is the

hydrogen bond carboxyl-carboxylate pairs establish at low pH but not at alkaline conditions, why substitution with an amide can enable this at high pH. The hydrogen bonding stabilizes the enzyme, leading to higher stability and activity. The results of this study deduce that favorable stabilizing interactions (in this case in alkaline conditions) that leads to improved activity at high pH and additionally that these are additive. As expected, the substitutions hampered the enzyme at lower pH, as the enzyme was not further stabilized under such condition, supporting the molecular understanding of the mechanism (Wohlfahrt et al., 2003). Another successful attempt to improve the performance has been done on the basis of multiple sequence alignment of GH6s with known crystal structures. The alignment found a correlation between a low number of charged residues (especially for the positively charged) and a low pH optimum. This theory was subjected to the endoglucanase CenA where several variants with different amounts of charged amino acids was substituted with alanine. They were found to gain activity at lower pH compared to the wild type with improved performance of up to 62 % (Cockburn & Clarke, 2011).

4.5.1 Altering the pH-temperature optimum by loop swap

Returning to the previously described loop deletion variant, dB2-1, that had its properties shifted from a cellobiohydrolase towards an endoglucanase. This variant was strongly modified by deletion of 14 amino acids (Figure 31), and one could wonder how this enzyme is affected by pH and temperature. The following data is unpublished, however the enzyme is described in **paper I**, and the method used is equivalent to the one described in **paper II**.

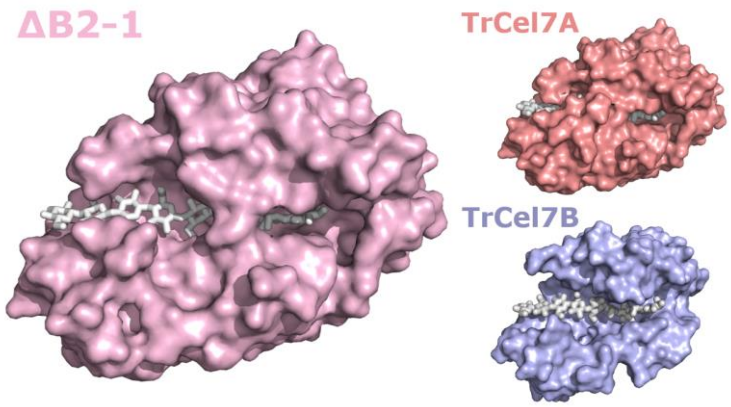
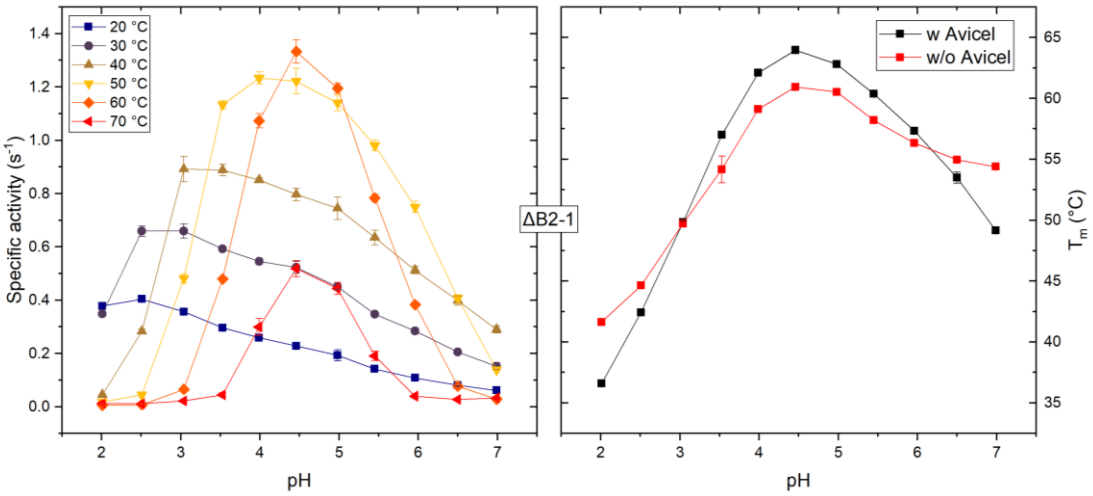


Figure 31. A putative representation of the surface structure of ΔB2-1 (missing the B2 loop compared to TrCel7A). The homology model was created using Phyre2 (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015) based on Cel7A PDB entry "4C4C" using default setting. On the right is small representations of TrCel7A (PDB 4C4C) and TrCel7B (PBD 1EG1).

Recall the horizontal curves from pH 2 - 5.5 at 20 - 50 °C of TrCel7A and the TrCel7B which displayed bell-shaped curves during all conditions (recapped in Figure 32). Once again, we see how the results of ΔB2-1 are shifted from the TrCel7A towards the TrCel7B (Figure 32). This applies to increased specific activity, and that the enzyme is denatured at the highest tested temperature, 70 °C. We also see, that suddenly the enzyme is dependent on pH at all temperatures, hence also during stabile conditions (20 °C and 30 °C). The catalytic optimum is markedly shifted down to pH 2.5 (20 °C), but as temperature is increased, the optimum shifts systematically from pH 2.5 towards the stability optimum, which remained pH 4.5 (Figure 32, right). This again stress that the stability optimum has large influence on the catalytically efficiency of an enzyme. The stability data of this variant, ΔB2-1, shows how the substrate do not stabilize the enzyme at all pH values, as the curve with and without Avicel crosses over. This might suggest that some interaction of the complexation has been impaired. One obvious interaction removed, is the W192 that interacts with the Y51 of the B1 loop that probably keeps both loops stabilized. In addition, the T_m, as also presented in **paper I**, was lowered compared to the wild type. Altogether, this implies that the coverage of the tunnel is important for the pH-dependency of the TrCel7A, because of the accessibility of water into the tunnel. As soon as the B2 loop is removed, we see how the parameters change towards those of an EG. Summing up, this emphasize how the structure influence kinetics and pH stability, and it might be due to the openness and the accessibility to the binding area.

A



B

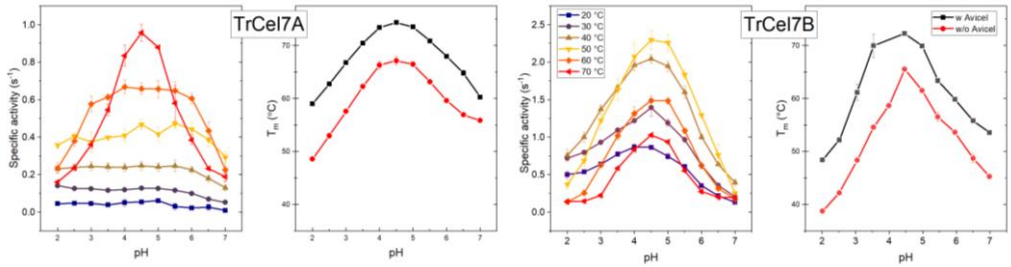


Figure 32. Specific activity (90 g/L Avicel, 100 nM enzyme) at different temperatures (20 – 70 C°) (A, left) and T_m with and without 90 g/L Avicel, 1 µM enzyme (A, right) for TrCel7A ΔB2-1 as function of pH. These data are unpublished but performed similarly to **paper II**. Below are recap of data for TrCel7A and TrCel7B (equivalent to Figure X).

Altogether, the finding of this chapter is relevant news from an industrial point of view. We found that TrCel7A is more tolerant to low pH, while hydrolyzing cellulose, than previously known. Since acid from the pretreatment can be released from the biomass during enzymatic hydrolysis (Mosier et al., 2005), it is very convenient that the main workhorse of the enzymatic cocktail is robust during deconstruction of the biomass. We also found, that one should be careful using small soluble substrate analogs to reflect the characteristics of cellulose hydrolysis. In addition, the increased T_m during complexation with substrate is also positive, as this prevents the enzyme from denaturing during long term hydrolysis. These results can contribute to future screening and investigations of cellulase performance for many different industrial applications.

5 Concluding remarks

The foundation of this thesis is based on comparative biochemical characterization of cellulases to shed light on mechanistic aspects. The findings builds partly on the application of a simple but stringent framework of conventional and inverse Michaelis-Menten kinetics and partly on a systematic application of identical materials/substrates and methods. In addition, it builds on a consistent use of the same enzyme expression host and thereby uniform posttranslational modifications. The subject is focused on investigations revolving around the function and structure of the cellobiohydrolase (CBH) of *T. reesei*, TrCel7A, as well as genetically modified variants, using different well-established and new experimental methods, with a special focus on how the loops affects several aspects of the enzyme.

The Cel7A plays a central role of natural and industrial cellulose degradation, and in order to improve it for cost-effective industrial utilization, it is crucial to study and understand its functions and engineering opportunities. By utilizing the enzymes intrinsic fluorescence, we developed a method, where we could measure the threading of TrCel7A with a cellulosic substrate in real-time. This technique made us able to measure the dynamics of a library of interesting variants, including singularly mutated tunnel tryptophans into alanine. Results showed that threading was much faster than dissociation and that the highly conserved tryptophans hardly affected the rate of complexation between enzyme and substrate.

The loops are known to be responsible for a wide array of properties and functions and is also subject for mutations that alter activity and binding significantly. We investigated the loops unique to TrCel7A (CBH) compared to TrCel7B (EG) by creating a library of variants with deletion of different loops followed by biochemical and kinetic characterization. We found that the B2 loop is the key determinant to the CBH-like behavior as we found several properties shifting from CBH towards EG-like behavior after deletion. A closer look on the B2 loop, investigated by site saturation of the conserved N200, revealed how we can engineer the cellulase to have a higher and lower maximum speed, however, we were not able improve the amount of attack sites. This suggest that the conserved asparagine is evolutionary optimized for exactly this, and not for instance to gain the highest speed possible.

Finally, we investigated how pH affects stability and activity at several different temperatures and how the activity is dependent on the size of the substrate being hydrolyzed. The long catalytic binding tunnel of TrCel7A can withstand very low pH as long as cellulose is the substrate and the temperature is kept below the stabilizing

conditions. During hydrolysis of a small substrate analog, this is not the case, and a classic bell shaped curve arise with apex around pH 4.5 reflecting most of the literature. However, these findings were in contrast to TrCel7B that mostly differs from Cel7A by its open binding cleft and display usual pH dependency, as we also experienced with the variant lacking the B2 loop. As the truncating of loops are the major difference between the two, we can again speculate that the majority of the observed effect is due to these.

Loops are flexible and sequentially one of the places where enzymes varies the most. With this work, we have contributed to the understanding of the properties of these as well as gained insights in how engineering of loops can alter several biochemical and kinetic parameters significantly. We elucidated how cellulases work during several different pH and temperature conditions and found how the pH dependency also is influenced by the loops and structure as well as investigated the rate of complexation. Together, we gained further insights that might contribute to future engineering for improvement of this industrially relevant machinery.

6 References

Andersen, M., Kari, J., Borch, K., & Westh, P. (2018). Michaelis–Menten equation for degradation of insoluble substrate. *Mathematical biosciences, 296*, 93-97.

Bailey, C. J. (1989). Enzyme kinetics of cellulose hydrolysis. *Biochemical Journal, 262*(3), 1001.

Bansal, P., Hall, M., Realff, M. J., Lee, J. H., & Bommarius, A. S. (2009). Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnology advances, 27*(6), 833-848.

Becker, D., Braet, C., Brumer, H., Claeyssens, M., Divne, C., Fagerström, B. R., . . . Koivula, A. (2001). Engineering of a glycosidase Family 7 cellobiohydrolase to more alkaline pH optimum: the pH behaviour of *Trichoderma reesei* Cel7A and its E223S/A224H/L225V/T226A/D262G mutant. *Biochemical Journal, 356*(1), 19-30. doi:10.1042/0264-6021:3560019

Berg, J. M., Tymoczko, J. L., & Stryer, L. (2012). *Biochemistry*. New York: WH Freeman.

Bezerra, R. M., & Dias, A. A. (2004). Discrimination among eight modified Michaelis-Menten kinetics models of cellulose hydrolysis with a large range of substrate/enzyme ratios. *Applied biochemistry and biotechnology, 112*(3), 173-184.

Bhat, M. (2000). Cellulases and related enzymes in biotechnology. *Biotechnology advances, 18*(5), 355-383.

Boer, H., & Koivula, A. (2003). The relationship between thermal stability and pH optimum studied with wild-type and mutant *Trichoderma reesei* cellobiohydrolase Cel7A. *European journal of biochemistry, 270*(5), 841-848.

Borisova, A. S., Eneyskaya, E. V., Bobrov, K. S., Jana, S., Logachev, A., Polev, D. E., . . . Sandgren, M. (2015). Sequencing, biochemical characterization, crystal structure and molecular dynamics of cellobiohydrolase Cel7A from *Geotrichum candidum* 3C. *The FEBS journal, 282*(23), 4515-4537. doi:10.1111/febs.13509

Borisova, A. S., Eneyskaya, E. V., Jana, S., Badino, S. F., Kari, J., Amore, A., . . . Himmel, M. E. (2018). Correlation of structure, function and protein dynamics in GH7 cellobiohydrolases from *Trichoderma atroviride*, *T. reesei* and *T. harzianum*. *Biotechnology for biofuels, 11*(1), 5. doi:10.1186/s13068-017-1006-7

Bu, L., Crowley, M. F., Himmel, M. E., & Beckham, G. T. (2013). Computational investigation of pH dependence on loop flexibility and catalytic function in glycoside hydrolases. *Journal of Biological Chemistry*, jbc. M113. 462465.

Bu, L., Nimlos, M. R., Shirts, M. R., Ståhlberg, J., Himmel, M. E., Crowley, M. F., & Beckham, G. T. (2012). Product binding varies dramatically between processive and nonprocessive cellulase enzymes. *Journal of Biological Chemistry, 287*(29), 24807-24813.

Carriquiry, M. A., Du, X., & Timilsina, G. R. (2011). Second generation biofuels: Economics and policies. *Energy Policy, 39*(7), 4222-4234.

Cockburn, D. W., & Clarke, A. J. (2011). Modulating the pH-activity profile of cellulase A from *Cellulomonas fimi* by replacement of surface residues. *Protein Engineering, Design & Selection, 24*(5), 429-437.

Colussi, F., Garcia, W., Rosseto, F. R., de Mello, B. L. S., de Oliveira Neto, M., & Polikarpov, I. (2012). Effect of pH and temperature on the global compactness, structure, and activity of cellobiohydrolase Cel7A from *Trichoderma harzianum*. *European Biophysics Journal, 41*(1), 89-98.

Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Tatsumi, H., Spodsberg, N., Borch, K., & Westh, P. (2012). Pre-steady state kinetics for the hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A. *Journal of Biological Chemistry*, jbc. M111. 334946.

Cruys-Bagger, N., Tatsumi, H., Ren, G. R., Borch, K., & Westh, P. (2013). Transient kinetics and rate-limiting steps for the processive cellobiohydrolase Cel7A: effects of substrate structure and carbohydrate binding domain. *Biochemistry, 52*(49), 8938-8948.

Cruys-Bagger, N., Elmerdahl, J., Praestgaard, E., Borch, K., & Westh, P. (2013). A steady-state theory for processive cellulases. *The FEBS journal, 280*(16), 3952-3961.

Cruys-Bagger, N., Ren, G., Tatsumi, H., Baumann, M. J., Spodsberg, N., Andersen, H. D., . . . Westh, P. (2012). An amperometric enzyme biosensor for real-time measurements of cellobiohydrolase activity on insoluble cellulose. *Biotechnology and bioengineering, 109*(12), 3199-3204.

Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure, 3*(9), 853-859. doi:https://doi.org/10.1016/S0969-2126(01)00220-9

Davison, B. H., Parks, J., Davis, M. F., & Donohoe, B. S. (2013). Plant cell walls: basics of structure, chemistry, accessibility and the influence on conversion. In C. E. Wyman (Ed.), *Aqueous pretreatment of plant biomass for biological and chemical conversion to fuels and chemicals* (pp. 23-38): John Wiley & Sons, Ltd.

Desprez, T., Juraniec, M., Crowell, E. F., Jouy, H., Pochylova, Z., Parcy, F., . . . Vernhettes, S. (2007). Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, 104*(39), 15572-15577.

Divne, C., Ståhlberg, J., Teeri, T. T., & Jones, T. A. (1998). High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *Journal of molecular biology, 275*(2), 309-325. doi:10.1006/jmbi.1997.1437

Duchesne, L. C., & Larson, D. (1989). Cellulose and the evolution of plant life. *Bioscience, 39*(4), 238-241.

Fersht, A. (1999). *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*: Macmillan.

GhattyVenkataKrishna, P. K., Alekozai, E. M., Beckham, G. T., Schulz, R., Crowley, M. F., Uberbacher, E. C., & Cheng, X. (2013). Initial recognition of a cellodextrin chain in the cellulose-binding tunnel may affect cellobiohydrolase directional specificity. *Biophysical journal, 104*(4), 904-912.

Gregg, D., Boussaid, A., & Saddler, J. (1998). Techno-economic evaluations of a generic wood-to-ethanol process: effect of increased cellulose yields and enzyme recycle. *Bioresource technology, 63*(1), 7-12.

Hall, M., Bansal, P., Lee, J. H., Realff, M. J., & Bommarius, A. S. (2011). Biological pretreatment of cellulose: Enhancing enzymatic hydrolysis rate using cellulose-binding domains from cellulases. *Bioresource technology, 102*(3), 2910-2915.

Harris, T. K., & Turner, G. J. (2002). Structural basis of perturbed p*K*a values of catalytic groups in enzyme active sites. *IUBMB life, 53*(2), 85-98.

Hawksworth, D. (2011). A new dawn for the naming of fungi: impacts of decisions made in Melbourne in July 2011 on the future publication and regulation of fungal names. *MycoKeys, 1*, 7-20. doi:10.3897/mycokeys.1.2062

Heinzelman, P., Snow, C. D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., . . . Arnold, F. H. (2009). A family of thermostable fungal cellulases created by structure-guided recombination. *Proceedings of the National Academy of Sciences, 106*(14), 5610-5615.

Henrissat, B., Teeri, T., & Warren, R. (1998). A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *FEBS letters, 425*(2), 352-354.

Herlet, J., Kornberger, P., Roessler, B., Glanz, J., Schwarz, W., Liebl, W., & Zverlov, V. (2017). A new method to evaluate temperature vs. pH activity profiles for biotechnological relevant enzymes. *Biotechnology for biofuels, 10*(1), 234.

Himmel, M. E., Ruth, M. F., & Wyman, C. E. (1999). Cellulase for commodity products from cellulosic biomass. *Current opinion in biotechnology, 10*(4), 358-364.

Hobdey, S. E., Knott, B. C., Momeni, M. H., Taylor, L. E., Borisova, A. S., Podkaminer, K. K., . . . Beckham, G. T. (2016). Biochemical and structural characterization of two *Dictyostelium* cellobiohydrolases from the Amoebozoa kingdom reveal a high conservation between distant phylogenetic trees of life. *Applied and environmental microbiology*, AEM. 00163-00116. doi:10.1128/aem.00163-16

Igarashi, K., Koivula, A., Wada, M., Kimura, S., Penttilä, M., & Samejima, M. (2009). High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. *The Journal of biological chemistry, 284*(52), 36186-36190. doi:10.1074/jbc.M109.034611

IPCC, I. P. o. C. C. (2018). *Global carbon budget 2018.* Retrieved from www.globalcarbonproject.org/carbonbudget/18/files/GCP_CarbonBudget_2018.pdf

Jalak, J., Kurašhin, M., Teugjas, H., & Väljamäe, P. (2012). Endo-exo synergism in cellulose hydrolysis revisited. *Journal of Biological Chemistry*, jbc. M112. 381624.

Jalak, J., & Väljamäe, P. (2010). Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis. *Biotechnology and bioengineering, 106*(6), 871-883. doi:10.1002/bit.22779

Jørgensen, H., Kristensen, J. B., & Felby, C. (2007). Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels, Bioproducts and Biorefining, 1*(2), 119-134.

Kari, J., Andersen, M., Borch, K., & Westh, P. (2017). An Inverse Michaelis–Menten Approach for Interfacial Enzyme Kinetics. *Acs Catalysis, 7*(7), 4904-4914.

Kari, J., Olsen, J., Borch, K., Cruys-Bagger, N., Jensen, K., & Westh, P. (2014). Kinetics of cellobiohydrolase (Cel7A) variants with lowered substrate affinity. *Journal of Biological Chemistry*, jbc. M114. 604264.

Karuna, N., & Jeoh, T. (2017). The productive cellulase binding capacity of cellulosic substrates. *Biotechnology and bioengineering, 114*(3), 533-542.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols, 10*(6), 845.

Kleman-Leyer, K. M., SiiKa-Aho, M., Teeri, T. T., & Kirk, T. K. (1996). The Cellulases Endoglucanase I and Cellobiohydrolase II of Trichoderma reesei Act Synergistically To Solubilize Native Cotton Cellulose but Not To Decrease Its Molecular Size. *Applied and environmental microbiology, 62*(8), 2883-2887.

Klemm, D., Heublein, B., Fink, H. P., & Bohn, A. (2005). Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte Chemie International Edition, 44*(22), 3358-3393.

Kleywegt, G. J., Zou, J.-Y., Divne, C., Davies, G. J., Sinning, I., Ståhlberg, J., . . . Jones, T. A. (1997). The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. *Journal of molecular biology, 272*(3), 383-397. doi:10.1006/jmbi.1997.1243

Knott, B. C., Haddad Momeni, M., Crowley, M. F., Mackenzie, L. F., Götz, A. W., Sandgren, M., . . . Beckham, G. T. (2013). The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. *Journal of the American Chemical Society, 136*(1), 321-329.

Knowles, J. K., Lentovaara, P., Murray, M., & Sinnott, M. L. (1988). Stereochemical course of the action of the cellobioside hydrolases I and II of *Trichoderma reesei*. *Journal of the Chemical Society, Chemical Communications*(21), 1401-1402.

Koivula, A., Ruohonen, L., Wohlfahrt, G., Reinikainen, T., Teeri, T. T., Piens, K., . . . Becker, D. (2002). The Active Site of Cellobiohydrolase Cel6A from *Trichoderma reesei*: The Roles of Aspartic Acids D221 and D175. *Journal of the American Chemical Society, 124*(34), 10015-10024.

Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., & Gronenborn, A. M. (1989). Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry, 28*(18), 7241-7257.

Kurašin, M., & Väljamäe, P. (2011). Processivity of cellobiohydrolases is limited by the substrate. *Journal of Biological Chemistry, 286*(1), 169-177.

Lai-Kee-Him, J., Chanzy, H., Müller, M., Putaux, J.-L., Imai, T., & Bulone, V. (2002). *In vitro* versus *in vivo* cellulose microfibrils from plant primary wall synthases: Structural differences. *Journal of Biological Chemistry, 277*(40), 36931-36939.

Lakowicz, J. R. (2013). *Principles of fluorescence spectroscopy, Chp 16* (3rd ed.): Springer Science & Business Media.

Lemos, M. A., Teixeira, J., Domingues, M., Mota, M., & Gama, F. (2003). The enhancement of the cellulolytic activity of cellobiohydrolase I and endoglucanase by the addition of cellulose binding domains derived from *Trichoderma reesei*. *Enzyme and microbial technology, 32*(1), 35-40.

Linder, M., Mattinen, M. L., Kontteli, M., Lindeberg, G., Ståhlberg, J., Drakenberg, T., . . . Annila, A. (1995). Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. *Protein science, 4*(6), 1056-1064.

Linder, M., & Teeri, T. T. (1997). The roles and function of cellulose-binding domains. *Journal of biotechnology, 57*(1-3), 15-28. doi:10.1016/S0168-1656(97)00087-4

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2013). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research, 42*(D1), D490-D495.

Lynd, L. R., Weimer, P. J., Van Zyl, W. H., & Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev., 66*(3), 506-577.

Marjamaa, K., Toth, K., Bromann, P. A., Szakacs, G., & Kruus, K. (2013). Novel *Penicillium* cellulases for total hydrolysis of lignocellulosics. *Enzyme and microbial technology, 52*(6-7), 358-369. doi:10.1016/j.enzmictec.2013.03.003

Martin-Martinez, F. J. (2018). Designing nanocellulose materials from the molecular scale. *Proceedings of the National Academy of Sciences, 115*(28), 7174-7175.

Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., . . . Cullen, D. (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature biotechnology, 26*(5), 553.

Matthews, J. F., Skopec, C. E., Mason, P. E., Zuccato, P., Torget, R. W., Sugiyama, J., . . . Brady, J. W. (2006). Computer simulation studies of microcrystalline cellulose Iβ. *Carbohydrate research, 341*(1), 138-152.

McLaren, A. D., & Packer, L. (1970). Some aspects of enzyme reactions in heterogeneous systems. *Adv Enzymol Relat Areas Mol Biol, 33*, 245-308.

Meinke, A., Damude, H. G., Tomme, P., Kwan, E., Kilburn, D. G., Miller, R. C., . . . Gilkes, N. R. (1995). Enhancement of the endo-β-1, 4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop. *Journal of Biological Chemistry, 270*(9), 4383-4386.

Momeni, M. H., Payne, C. M., Hansson, H., Mikkelsen, N. E., Svedberg, J., Engström, Å., . . . Ståhlberg, J. (2013). Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus *Heterobasidion irregulare*. *Journal of Biological Chemistry*, jbc. M112. 440891.

Morgan, J. L., Strumillo, J., & Zimmer, J. (2013). Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature, 493*(7431), 181.

Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapple, M., & Ladisch, M. (2005). Features of promising technologies for pretreatment of lignocellulosic

biomass. *Bioresource technology, 96*(6), 673-686. doi:10.1016/j.biortech.2004.06.025

Murphy, L., Bohlin, C., Baumann, M. J., Olsen, S. N., Sørensen, T. H., Anderson, L., . . . Westh, P. (2013). Product inhibition of five *Hypocrea jecorina* cellulases. *Enzyme and microbial technology, 52*(3), 163-169.

Nakamura, A., Tsukada, T., Auer, S., Furuta, T., Wada, M., Koivula, A., . . . Samejima, M. (2013). Tryptophan residue at active-site tunnel entrance of *Trichoderma reesei* cellobiohydrolase Cel7A is important to initiate degradation of crystalline cellulose. *Journal of Biological Chemistry*, jbc. M113. 452623.

Nelson, D. L., Lehninger, A. L., & Cox, M. M. (2008). *Lehninger principles of biochemistry*: Macmillan.

O'sullivan, A. C. (1997). Cellulose: the structure slowly unravels. *Cellulose, 4*(3), 173-207.

Ogunmolu, F. E., Jagadeesha, N. B. K., Kumar, R., Kumar, P., Gupta, D., & Yazdani, S. S. (2017). Comparative insights into the saccharification potentials of a relatively unexplored but robust *Penicillium funiculosum* glycoside hydrolase 7 cellobiohydrolase. *Biotechnology for biofuels, 10*(1), 71.

Olsen, H. S., & Falholt, P. (1998). The role of enzymes in modern detergency. *Journal of Surfactants and Detergents, 1*(4), 555-567. doi:10.1007/s11743-998-0058-7

Olsen, J. P., Alasepp, K., Kari, J., Cruys-Bagger, N., Borch, K., & Westh, P. (2016). Mechanism of product inhibition for cellobiohydrolase Cel7A during hydrolysis of insoluble cellulose. *Biotechnology and bioengineering, 113*(6), 1178-1186.

Olsen, J. P., Kari, J., Borch, K., & Westh, P. (2017). A quenched-flow system for measuring heterogeneous enzyme kinetics with sub-second time resolution. *Enzyme and microbial technology, 105*, 45-50.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein science, 4*(11), 2411-2423.

Pauly, M., & Keegstra, K. (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. *The Plant Journal, 54*(4), 559-568. doi:10.1111/j.1365-313X.2008.03463.x

Payne, C. M., Himmel, M. E., Crowley, M. F., & Beckham, G. T. (2011). Decrystallization of oligosaccharides from the cellulose Iβ surface with molecular simulation. *The Journal of Physical Chemistry Letters, 2*(13), 1546-1550.

Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., . . . Beckham, G. T. (2015). Fungal cellulases. *Chemical reviews, 115*(3), 1308-1448.

Pellegrini, V. O., Lei, N., Kyasaram, M., Olsen, J. P., Badino, S. F., Windahl, M. S., . . . Westh, P. (2014). Reversibility of substrate adsorption for the cellulases Cel7A, Cel6A, and Cel7B from *Hypocrea jecorina*. *Langmuir, 30*(42), 12602-12609.

Pellegrini, V. O., Serpa, V. I., Godoy, A. S., Camilo, C. M., Bernardes, A., Rezende, C. A., . . . Polikarpov, I. (2015). Recombinant *Trichoderma harzianum* endoglucanase I (Cel7B) is a highly acidic and promiscuous carbohydrate-active enzyme. *Applied microbiology and biotechnology, 99*(22), 9591-9604.

Penttilä, M., Lehtovaara, P., Nevalainen, H., Bhikhabhai, R., & Knowles, J. (1986). Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the endoglucanase I gene. *Gene, 45*(3), 253-263.

Petrauskas, V., Zubrienė, A., Todd, M. J., & Matulis, D. (2019). Inhibitor Binding to Carbonic Anhydrases by Fluorescent Thermal Shift Assay. In *Carbonic Anhydrase as Drug Target* (pp. 63-78): Springer.

Praestgaard, E., Elmerdahl, J., Murphy, L., Nymand, S., McFarland, K., Borch, K., & Westh, P. (2011). A kinetic model for the burst phase of processive cellulases. *The FEBS journal, 278*(9), 1547-1560.

Quinlan, R. J., Sweeney, M. D., Leggio, L. L., Otten, H., Poulsen, J.-C. N., Johansen, K. S., . . . Anthonsen, A. (2011). Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proceedings of the National Academy of Sciences, 108*(37), 15079-15084.

Reese, E. (1976). *History of the cellulase program at the US army Natick Development Center*. Paper presented at the Biotechnol. Bioeng. Symp.;(United States).

Reier, G. E. (2000). Avicel PH Microcrystalline Cellulose, NF, Ph Eur., JP, BP. *FMC Corporation, 11*, 1-27.

Rosgaard, L., Pedersen, S., Langston, J., Akerhielm, D., Cherry, J. R., & Meyer, A. S. (2007). Evaluation of minimal *Trichoderma reesei* cellulase mixtures on differently pretreated barley straw substrates. *Biotechnology progress, 23*(6), 1270-1276.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J., & Jones, T. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science, 249*(4967), 380-386.

Schülein, M. (1997). Enzymatic properties of cellulases from *Humicola insolens*. *Journal of biotechnology, 57*(1-3), 71-81. doi:10.1016/S0168-1656(97)00090-4

Silveira, R. L., & Skaf, M. S. (2018). Concerted motions and large-scale structural fluctuations of Trichoderma reesei Cel7A cellobiohydrolase. *Physical Chemistry Chemical Physics, 20*(11), 7498-7507.

Smits, J., Rinzema, A., Tramper, J., Van Sonsbeek, H., Hage, J., Kaynak, A., & Knol, W. (1998). The influence of temperature on kinetics in solid-state fermentation. *Enzyme and microbial technology, 22*(1), 50-57.

Sørensen, T. H., Cruys-Bagger, N., Windahl, M. S., Badino, S. F., Borch, K., & Westh, P. (2015). Temperature effects on kinetic parameters and substrate affinity of Cel7A cellobiohydrolases. *Journal of Biological Chemistry*, jbc. M115. 658930.

Sørensen, T. H., Windahl, M. S., McBrayer, B., Kari, J., Olsen, J. P., Borch, K., & Westh, P. (2017). Loop variants of the thermophile *Rasamsonia emersonii* Cel7A with improved activity against cellulose. *Biotechnology and bioengineering, 114*(1), 53-62. doi:10.1002/bit.26050

Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeyssens, M., Teeri, T. T., & Jones, T. A. (1996). Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from*Trichoderma reesei*. *Journal of molecular biology, 264*(2), 337-349.

Ståhlberg, J., Henriksson, H., Divne, C., Isaksson, R., Pettersson, G., Johansson, G., & Jones, T. A. (2001). Structural basis for enantiomer binding and separation of a common β-blocker: crystal structure of cellobiohydrolase Cel7A with bound (S)-propranolol at 1.9 Å resolution1. *Journal of molecular biology, 305*(1), 79-93.

Ståhlberg, J., Johansson, G., & Pettersson, G. (1993). *Trichoderma reesei* has no true exo-cellulase: all intact and truncated cellulases produce new reducing end groups on cellulose. *Biochimica et Biophysica Acta (BBA)-General Subjects, 1157*(1), 107-113.

Taylor, C. B., Payne, C. M., Himmel, M. E., Crowley, M. F., McCabe, C., & Beckham, G. T. (2013). Binding site dynamics and aromatic–carbohydrate interactions in processive and non-processive family 7 glycoside hydrolases. *The Journal of Physical Chemistry B, 117*(17), 4924-4933.

Taylor, L. E., Knott, B. C., Baker, J. O., Alahuhta, P. M., Hobdey, S. E., Linger, J. G., . . . Podkaminer, K. (2018). Engineering enhanced cellobiohydrolase activity. *Nature communications, 9*(1), 1186.

Teugjas, H., & Väljamäe, P. (2013). Product inhibition of cellulases studied with 14 C-labeled cellulose substrates. *Biotechnology for biofuels, 6*(1), 104.

Texier, H., Dumon, C., Neugnot-Roux, V., Maestracci, M., & O'Donohue, M. J. (2012). Redefining XynA from *Penicillium funiculosum* IMI 378536 as a GH7 cellobiohydrolase. *Journal of industrial microbiology & biotechnology, 39*(11), 1569-1576.

United-Nations. Retrieved from https://sustainabledevelopment.un.org/

Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørlie, M., & Eijsink, V. G. (2010). An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science, 330*(6001), 219-222.

Väljamäe, P., Sild, V., Nutt, A., Pettersson, G., & Johansson, G. (1999). Acid hydrolysis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I. *European journal of biochemistry, 266*(2), 327-334.

Várnai, A., Siika-aho, M., & Viikari, L. (2013). Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. *Biotechnology for biofuels, 6*(1), 30.

Von Ossowski, I., Ståhlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., . . . Mahdi, S. (2003). Engineering the exo-loop of *Trichoderma reesei* cellobiohydrolase, Cel7A. A comparison with *Phanerochaete chrysosporium* Cel7D. *Journal of molecular biology, 333*(4), 817-829.

Wang, T., Liu, X., Yu, Q., Zhang, X., Qu, Y., Gao, P., & Wang, T. (2005). Directed evolution for engineering pH profile of endoglucanase III from *Trichoderma reesei*. *Biomolecular engineering, 22*(1-3), 89-94.

Wingren, A., Galbe, M., & Zacchi, G. (2003). Techno-economic evaluation of producing ethanol from softwood: Comparison of SSF and SHF and identification of bottlenecks. *Biotechnology progress, 19*(4), 1109-1117. doi:10.1021/bp0340180

Wohlfahrt, G., Pellikka, T., Boer, H., Teeri, T. T., & Koivula, A. (2003). Probing pH-dependent functional elements in proteins: modification of carboxylic acid pairs

in *Trichoderma reesei* cellobiohydrolase Cel6A. *Biochemistry, 42*(34), 10095-10103.

Wolfenden, R., Lu, X., & Young, G. (1998). Spontaneous hydrolysis of glycosides. *Journal of the American Chemical Society, 120*(27), 6814-6815.

Zechel, D. L., & Withers, S. G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Accounts of chemical research, 33*(1), 11-18.

Zhang, Y. H., Cui, J., Lynd, L. R., & Kuang, L. R. (2006). A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules, 7*(2), 644-648.

Zhang, Y. H., & Lynd, L. R. (2004). Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnology and bioengineering, 88*(7), 797-824.